Coding Theory

notes and projections from lecture

Kit Tyabandha, PhD

God's Ayudhya's Defence
Bangkok

14^{th} January, 2007

Catalogue in Publication Data
Kit Tyabandha
Coding Theory, notes and projections from lecture: – Bangkok, Kittix, GAD, 2005
271 p.
1. Coding Theory I. Tyabandha, Kit II. Mathematics.
510
ISBN 974–94279–4–7

© Kit Tyabandha, 2005
All rights reserved

Published by Kittix Publishing
God's Ȧyudhyạ's Defence
1564/11 Prajarastrasaya 1 Road
Bangkok 10800, Thailand

Editor
Vaen Sriwayudhya

Printed in Thailand by
Kittix Press

Typeset using TEX

God's Ȧyudhyạ's Defence
and
Kittix

are the only two trademarks
relevant to the publication of this book

All other trademarks and trade names
are mentioned solely for explanation

In Thailand, Baht 1000
Elsewhere £20

To God.

Preface

This work began in October 2005 when I started teaching Coding Theory. Coding and cryptography are similar, but the main concern of the former is in the existence of a noisy environment whereas that of the latter is in the secrecy of the message from unintended party. The concept of entropy, being fundamental for the purposes of economy, privacy and reliability, play a role in all branches of both subjects.

In this course the main theme is coding theory. Cryptography was only mentioned briefly towards the end. As the students were fairly familiar with algebra, but had familiarity with neither finite fields nor polynomial rings, we had some practice sessions where the students tried their hands on problems. In all we had two practices, on 6 and 13 January, three quizzes, on 20 January, 3 and 10 February, and one midterm exam on 27 January 2006. Our final exam was on 23 February 2006.

The projections were adapted from the hand-outs given to students for the lecture. These hand-outs form here the notes from lecture. Both the notes and the projections are written on plain TEX. I stopped making projection after the lecture on Linear Code on 9 December 2005. The reason was because the nature of activities we did in class had changed. We spent a fair amount of our time doing the exercises and problems, so the lecturing was shortened. And since the students were by now more familiar with the subject, I needed only guide them through the hand-outs. Another reason was because I felt that I had been producing too many of them, so I did not want to waste more paper. To do a similar thing for other subjects in the future it would probably do well to limit the number of these projections to under 20 pages for each lecture.

I had been teaching at Mahidol University. But now I have been told by the heads of department of mathematics there that they are going to fire me for, for one thing, I worked too hard, and for another I always teach in English. To counter that, English *is* my first language, and I work for God, therefore I work day and night with minimum amount of sleep for Him readily and quite happily. I can see no reasons why this should make anybody unhappy, I do not even believe in politics I only believe in God.

I thank my students for allowing me the privilege of teaching them. I hope they have learnt from me as much as I have from them.

Kit Tyabandha, PhD
Bangkok, 14^{th} January, 2007

Table of Contents

a. List of Algorithms vi
b. List of Axioms vi
c. List of Corollaries vi
d. List of Definitions vi
e. List of Equations ix
f. List of Examples ix
g. List of Exercises x
h. List of Figures x
i. List of Lectures xi
j. List of Notes xii
k. List of Problems xii
l. List of Propositions xii
m. List of Tables xii
n. List of Theorems xiii

1. Error and distance 1
2. Exercise for error and distance 7
3. Entropy and mutual information 9
4. Group, field and finite field 21
5. Exercise for group, field and finite field 30
6. Bounds in coding 31
7. Group, polynomial and Hamming codes 41
8. Example for group, polynomial and Hamming codes 51
9. Exercise for group, polynomial and Hamming codes 53
10. Finite field- and BCH codes 54
11. Example for finite field- and BCH codes 66
12. Linear codes 68
13. Example for linear codes 72
14. Cyclic codes 75
15. Example for cyclic codes 77
16. Exercise for cyclic codes 80
17. Goppa codes 81
18. Exercise for Goppa codes 86
19. MDS code 88
20. Example for MDS codes 91
21. Cryptography 93
22. Course outline 229
23. Quiz 1 230
24. Midterm examination 233
25. Quiz 2 238
26. Quiz 3 240
27. Final examination 242
28. Students' scores 244

List of Algorithms

1. decoding algorithm 2
2. Gilbert bound 39
3. syndrome decoding algorithm 45
4. Hamming codes 50
5. Procedure for correcting up to t errors in BCH codes 65
6. ROT13 algorithm 94

List of Axioms

1. entropy 10
2. entropy 10
3. grouping 12
4. continuity of entropy 12

List of Corollaries

4[1]. distance of a u-error-detecting code 6
5[1]. v-error-correcting code 6
9[1]. Fano's inequality 15
23[1]. finite fields 28
47[1]. Hamming as special case 59
54[1]. dimension and size of a code 69
55[1]. 69
57[1]. q even 70
83[1]. 88
88[1]. 89

List of Definitions

1. code 1
2. communication channel 1
3. memoryless channel 2
4. symmetric channel 2
5. maximum likelihood decoding 2
6. Hamming distance 3
7. minimum distance decoding 4
8. distance of a code 5
9. error vector 5
10. detected and undetected errors 5
11. u-error-detecting code 5
12. v-error-correcting code 6
13. probability space and expectation 9

14. conditional probability 9
15. Markov chain 9
16. convexity 9
17. convex hull 9
18. convex cup and cap 10
19. Jensen's inequality 10
20. entropy 11
21. conditional entropy 14
22. mutual information 18
23. mutual information for three random variables 19
24. group 21
25. ring 21
26. regular- and singular elements 21
27. field 22
28. modulo 22
29. set of integer modulo m 24
30. notation for rings 24
31. characteristic of a field 24
32. subfield 25
33. polynomial ring 25
34. remainder of polynomial ring 26
35. the greatest common divisor and the least common multiple 26
36. finite field 29
37. primitive element 29
38. order 29
39. prime power 31
40. linear code 31
41. vector space 31
42. linearly independence 32
43. generating set and basis 32
44. relative minimum distance 32
45. optimal code 32
46. Hamming sphere 32
47. perfect code 33
48. asymptotic transmission rate 34
49. Group, ring and field 41
50. code words 41
51. distance function 41
52. weight function 41
53. homomorphism 42
54. group code 42
55. generator matrix 42
56. matrix code 42
57. parity check code 43
58. syndrome 45

59. dual codes 47
60. coset 47
61. vector space 47
62. linear-dependence 47
63. polynomial codes 48
64. matrix code 49
65. parity check matrix 49
66. coset 54
67. lcm 54
68. subring 54
69. ideal of a ring 55
70. congruence of a ring 55
71. coset of a ring 55
72. quotient ring 56
73. principal ideal 57
74. reducible polynomial 57
75. extension of a field 58
76. prime subfield 58
77. minimal polynomial 58
78. Vandermonde matrix 60
79. linear independence 68
80. linear span of a subspace 68
81. inner product of vectors 68
82. scalar product 68
83. basis 68
84. linear code 69
85. Hamming weight 70
86. elementary row operation 71
87. equivalent matrices 71
88. cyclic code 75
89. ideal of a ring 75
90. cyclic code 75
91. principal ideal 76
92. MDS code 81
93. trivial MDS 81
94. Reed-Solomon codes 81
95. generalised Reed-Solomon codes 82
96. alternant codes 82
97. Goppa codes 83
98. MDS code 88
99. 88
100. 89
101. encryption and decryption 93
102. key-based algorithms 93
103. cryptanalysis 93

104. security 94
105. substitution ciphers 94
106. transposition ciphers 94
107. one-time pad 94

List of Equations

1. probability of x given y 9
2. probability of y given x 9
3. convex cup 10
4. convex cap 10
5. Jensen's inequality for convex cup 10
6. Jensen's inequality for convex cap 10
7. Axiom 2 11
8. Starting from e 11
9. Put bounds on log's 11
10. Put the same bounds on h's 12
11. group of events 12
12. entropy for irrational probabilities 13
13. entropy 13
14. conditional entropy when y is given 14
15. conditional entropy 14
16. 34
17. 38
18. 39
19. 39
20. 39
21. 39
22. 39
23. 57
24. 57
25. 64
26. 64
27. 64
28. 64
29. 64
30. 64

List of Examples

1. codes 1
2. binary symmetric channel 2
3. the most likely word sent 2
4. two kinds of maximum likelihood decoding 3

5. Hamming distance 3
6. distance 4
7. Jensen's inequality in geometrical terms 10
8. grouping events 12
9. entropy of groups of events 13
10. scaling the entropy 13
11. units of entropy 13
12. inequality for a bound on $\ln x$ 15
13. inequality for a bound on $\ln \frac{1}{x}$ 16
14. two events 17
15. mutual information 18
16. mutual information in various forms 19
17. field 22
18. ring of integer modulo m 24
19. factorisation 26
20. greatest common divisor 26
21. rings of polynomial 27
22. another ring of polynomial 27
23. analogies between $\mathbf{Z}$ and $F[x]$ 27
24. prime power 31
25. linear binary code 31
26. vector space over finite field 31
27. Hamming bound for binary codes 34
28. finite field 41
29. generator and parity check matrices 43
30. 44
31. 57
32. ideal 75
33. ideal 76
34. example of substitution ciphers 94
35. example of substitution ciphers 94
36. example of a transposition cipher 94

List of Exercises

1. maximum likelihood decoding 3
2. IMLD 3
3. CMLD 3
4. MLD 3
5. ternary-code decoding 3
6. minimum distance decoding 5
7. nearest neighbour decoding 5
8. d dictates m 5
9. prove rings of polynomial theorem 27

List of Figures

1. code 1
2. Proof of triangular inequality of Hamming distance 4
3. inquality for $\ln x$ 16
4. $\log x$ 16
5. inequality for $\ln \frac{1}{x}$ 16
6. $\log \frac{1}{x}$ 17
7. entropy of two events 17
8. Distribution of students' Quiz 1 scores. 245
9. Distribution of students' Midterm scores. 246
10. Distribution of students' Quiz 2 scores. 247
11. Distribution of students' Quiz 3 scores. 249
12. Distribution of students' final exam scores. 250
13. Distribution of students' total score, Coding Theory 252

List of Lectures

1. Error and distance 1
2. Exercise for error and distance 7
3. Entropy and mutual information 9
4. Group, field and finite field 21
5. Exercise for group, field and finite field 30
6. Bounds in coding 31
7. Group, polynomial and Hamming codes 41
8. Example for group, polynomial and Hamming codes 51
9. Exercise for group, polynomial and Hamming codes 53
10. Finite field- and BCH codes 54
11. Example for finite field- and BCH codes 66
12. Linear codes 68
13. Example for linear codes 72
14. Cyclic codes 75
15. Example for cyclic codes 77
16. Exercise for cyclic codes 80
17. Goppa codes 81
18. Exercise for Goppa codes 86
19. MDS code 88
20. Example for MDS codes 91
21. Cryptography 93
22. Course outline 229
23. Quiz 1 230
24. Midterm examination 233
25. Quiz 2 238
26. Quiz 3 240
27. Final examination 242

28. Students' scores 244

List of Notes

1. dictionary's size 33
2. bounds on rate 34
3. double summations 36
4. Plotkin's bound 36
5. Gilbert bound 39
6. parity check matrix's property 44
7. assumptions for polynomial codes 48
8. Hamming codes 50
9. 54
10. 55
11. 55
12. addition and multiplication of congruences 55
13. 56
14. from Hamming to BCH code 59
15. 64
16. orthogonal complement 68
17. 75
18. 75

List of Problems

1. Stirling's approximation of $n!$ 33
2. dictionary size 34
3. 35
4. double summations 36
5. upper bound 37
6. 54
7. 56
8. 59
9. decoding BCH 63
10. 70
11. 88
12. 89
13. 90
14. 90

List of Propositions

1. Elias bound for binary alphabet 37

List of Tables

1. properties of a field 22
2. addition and multiplication tables of a ring of polynomial 27
3. addition and multiplication tables, another ring of polynomial 27
4. The analogies between $\mathbf{Z}$ and $F[x]$ 27
5. 77
6. Students' midterm scores. 244
7. Mark and rank of students' scores from Quiz 1. 244
8. Mark and rank of students' scores from Midterm Exam. 245
9. Mark and rank of students' scores from Midterm Exam. 245
10. Mark and rank of students' scores from Quiz 2. 246
11. Mark and rank of students' scores from 247
12. Scores and ranks of Quiz 3. 248
13. Scores and ranks of Final Exam. 249
14. Practice and attendance. 250
15. Total score, Coding Theory, second term, 2005–6 251
16. Total score and rank, Coding Theory, 2005–6 251
17. Grading schemes 252
18. Grades according to our grading scheme 253

List of Theorems

1. Hamming distance and the forward channel probability 3
2. triangular inequality of Hamming distance 4
3. maximum likelihood- and minimum distance decoding rules 5
4. distance of a u-error-detecting code 5
5. v-error-correcting code 6
6. entropy for equally likely events 11
7. entropy 13
8. conditional entropy 14
9. conditional entropy 14
10. maximum entropy 15
11. mutual information 18
12. mutual information for three random variables 19
13. Markov's chain 19
14. uniqueness of an identity 21
15. negative 22
16. zeros 22
17. congruence 23
18. properties of modulo 23
19. condition when $\mathbf{Z}_m$ is a field 24
20. characteristic of a field 24
21. elements of finite field 25
22. rings of polynomial 26

23. finite fields 28
24. linear code 31
25. vector space 32
26. Hamming bound 32
27. Plotkin's bound 35
28. weight and distance 41
29. monomorphic generator matrix 42
30. matrix code a group code 42
31. parity check code 43
32. minimum distance 43
33. parity check 45
34. generator and parity-check matrices 45
35. cosets 47
36. isomorphic vector spaces 48
37. divisible polynomial 48
38. minimum distance of a polynomial code 49
39. polynomial code a matrix code 49
40. 54
41. subring 55
42. 56
43. quotient ring 56
44. polynomial ring 57
45. polynomial ring a principal ideal ring 57
46. quotient ring a field 58
47. linear code correction capability 59
48. Vandermonde matrix 60
49. linearly independence 61
50. BCH code 61
51. minimum distance of linear code 63
52. Singleton bound 63
53. bound for BCH codes 63
54. dimension of a vector space 68
55. span and dual 69
56. double orthogonal 69
57. Hamming weight and distance 70
58. inequality 70
59. 75
60. MDS 81
61. BCH codes 81
62. RS codes are MDS 81
63. extended RS 82
64. generalised Reed-Solomon codes 82
65. generator matrix for RS code 82
66. dual of the generalised Reed-Solomon code 82
67. parity-check matrix of GRS 82

68. 83
69. dual of an alternant code....83
70. 83
71. Goppa....83
72. 83
73. Goppa code is alternant code....84
74. Goppa code is GRS code....84
75. 84
76. Goppa code is trace code....84
77. 84
78. 85
79. 88
80. 88
81. 88
82. 88
83. 88
84. 89
85. 89
86. 89
87. 89
88. 89
89. 90
90. 90

Error and distance

28^{th} October 2005

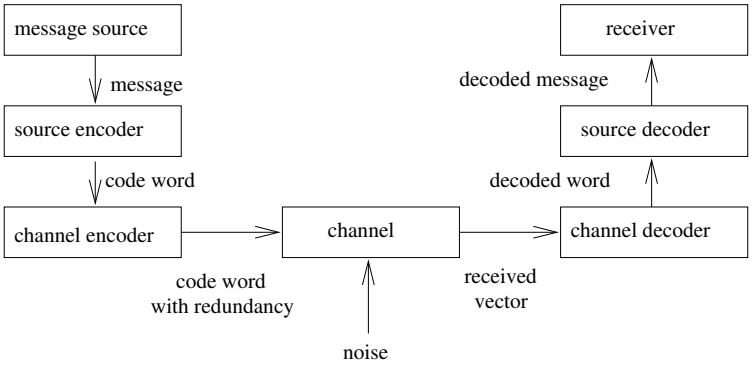


Figure 1 *Encoding and decoding of message. The channel encoder and decoder are there to introduce redundancy which let us detect and correct errors.*

Criteria for designing channel encoding algorithm and for the construction of the encoder and the decoder are namely fast encoding and decoding of messages, easy transmission of encoded messages, maximum rate of transfer of information, and maximum detection or correction capability.

Definition 1. Let $A = \{a_1, \ldots, a_q\}$ be a *code alphabet* of size q, and its elements are the *code symbols*. We call a *q-ary word* of length n over A a sequence $\mathbf{w} = w_1 \cdots w_n$, or equivalently a vector $(w_1, \ldots, w_n)$, where $w_i \in A$ for all i. We call a *q-ary block code* of length n over A a nonempty set C of q-ary words, that is *code words*, all of which is of the same length n. The number of code words C contains is the *size* m of C, consequently $m = |C|$. The *information rate* of the code C is $(\log_q |C|)/n$. We call an $(n, m) - code$ a code of length n and size m.

§

From Definition 1 we can see that C is a code containing code words each of which is composed of symbols from the code alphabet. A q-ary block code is a set of q-ary code words.

Example 1. A code over the code alphabet $\mathbf{F}_2 = \{0, 1\}$ is called a *binary code*, one over $\mathbf{F}_3 = \{0, 1, 2\}$ is called a *ternary code*. The term *quaternary code* refers to a code over either $\mathbf{F}_4 = \{0, 1, 2, 3\}$ or $\mathbf{Z}_4 = \{0, 1, 2, 3\}$.

Definition 2. A *communication channel* consists of a finite *channel alphabet* $A = \{a_1, \ldots, a_q\}$ together with a set of *forward channel probabilities* $p_{a_{ij}}$, such that for all i

$$\sum_{j=1}^{q} p_{a_{ij}} = 1$$

where $p_{a_i j}$ is the conditional probability that a_j is received, given that a_i is sent. If $\mathbf{x}$ is the word received when a word $\mathbf{c}$ was sent, e is the number of places where $\mathbf{x}$ and $\mathbf{c}$ differ, and n the length of each word, then the forward channel probability is $p_{\mathbf{cx}} = p^e(1-p)^{n-e}$.

§

The probability p_{ab} is normally written $p(b\,\text{received}|a\,\text{sent})$.

Definition 3. Let $\mathbf{c} = c_1 \cdots c_n$ and $\mathbf{x} = x_1 \cdots x_n$ be words of length n. Then a communication channel is said to be *memoryless* if

$$p_{\mathbf{cx}} = \prod_{i=1}^{n} p_{c_i x_i}$$

§

Definition 3 tells us that a communication channel is memoryless if the outcome of any one transmission is independent of the outcome of the previous transmissions.

Definition 4. A memory less channel with a channel alphabet of size q is called a *q-ary symmetric channel* if each symbol transmitted has the same probability $p < \frac{1}{2}$ of being received in error, and whenever a wrong symbol is received, each of the $q-1$ possible errors is equally likely. If $p > \frac{1}{2}$, the channel is known to be *useless.*

§

Example 2. The *binary symmetric channel* (BSC) is a memoryless channel having a channel alphabet $\{0,1\}$ and channel probabilities $p_{01} = p_{10} = p$ and $p_{00} = p_{11} = 1-p$. This probability of a bit error p in a BSC is called the *cross-over probability* of the BSC.

Example 3. When a received word is not among the vocabulary of the code, the most likely word sent is the one whose $p_{\mathbf{c}_i \mathbf{x}_i}$ is maximum over all $i = 1, \ldots, m$. A rule for finding the most likely code word sent in case of an error is called a *decoding rule.*

The procedure for finding the most likely message sent is described in Algorithm 1. Here $\mathbf{c}_i^x$ means the word deduced to be the actual code word sent to the best of our guess.

Algorithm 1 *Decoding algorithm*

for all words $\mathbf{x}_i$ received **do**
  if $\mathbf{x}_i$ *is not a valid code word* **then**
    $\mathbf{c}_i^x \leftarrow$*the most likely code word* $\mathbf{c}_i$ *sent, according to the decoding rule used*
  else
    $\mathbf{c}_i^x \leftarrow \mathbf{x}_i$
  endif
endfor

Definition 5. The *maximum likelihood decoding* is $p_{\mathbf{c}_i^x\mathbf{x}} = \max_{\mathbf{c}\in C} p_{\mathbf{c}\mathbf{x}}$, where $\mathbf{x}$ is the word received.

§

Example 4. Two kinds of maximum likelihood decoding are, when it happens that there are more than one word that has the same maximum likelihood, the *complete maximum likelihood decoding* chooses one of them arbitrarily, while the *incomplete maximum likelihood decoding* rejects all of them and asks for a retransmission.

§

Exercise 1. Code words from the binary code $\{00001, 00111, 02020, 00001\}$ are sent over a binary symmetric channel with the cross-over probability $p = 0.002$. Using the maximum likelihood decoding rule, decode the words, 01111, 01110, 11000, 10101 and 11111.

§

Exercise 2. Write the IMLD (incomplete maximum likelihood decoding) table for the code $C = \{001, 100, 110, 111\}$, and then again for the code $C = \{101, 111, 110\}$.

§

Exercise 3. Write the CMLD (complete maximum likelihood decoding) table for the code $C = \{110, 101, 011, 001, 100\}$, and then again for the code $C = \{000, 111, 010, 101\}$.

§

Exercise 4. A memoryless binary channel with channel probabilities $p_{00} = 0.81$ and $p_{11} = 0.95$. Code words from the code $C = \{000, 001, 011, 111\}$ are being sent over the channel. With the help of the maximum likelihood decoding rule, decode the words 010, 101, 100 and 110.

§

Exercise 5. A ternary code is $C = \{01202, 21201, 11220, 00112\}$. Using the nearest neighbour decoding rule, decode the words 01112, 02221, 12121 and 01012.

§

Definition 6. Let $\mathbf{x} = x_1 \cdots x_n$ and $\mathbf{y} = y_1 \cdots y_n$ be words of length n over an alphabet A. Then the *Hamming distance* between $\mathbf{x}$ and $\mathbf{y}$, denoted $\mathrm{d}(\mathbf{x}, \mathbf{y})$, is the number of places where $\mathbf{x}$ and $\mathbf{y}$ are different from each other, and

$$\mathrm{d}(\mathbf{x}, \mathbf{y}) = \mathrm{d}(x_1, y_1) + \ldots + \mathrm{d}(x_n, y_n)$$

where

$$\mathrm{d}(x_i, y_i) = \begin{cases} 0 & \text{if } x_i = y_i \\ 1 & \text{if } x_i \neq y_i \end{cases}$$

§

Theorem 1. The Hamming distance $\mathrm{d}(\mathbf{x}, \mathbf{c}) = i$ corresponds to the forward channel probability $p_{\mathbf{cx}} = p^i(1-p)^{n-i}$.

Proof. This is obvious from Definition's 2 and 6. □

Example 5. From Definition 6 it follows that $0 \leq \mathrm{d}(\mathbf{x}, \mathbf{y}) \leq n$; $\mathrm{d}(\mathbf{x}, \mathbf{y}) = 0$ if and only if $\mathbf{x} = \mathbf{y}$; and $\mathrm{d}(\mathbf{x}, \mathbf{y}) = \mathrm{d}(\mathbf{y}, \mathbf{x})$.

Example 6. Let A be the roman alphabet. If $\mathbf{x}$ = 'breed', $\mathbf{y}$ = 'bread', and $\mathbf{z}$ = 'break', then $\mathrm{d}(\mathbf{x}, \mathbf{y}) = \mathrm{d}(\mathbf{y}, \mathbf{z}) = 1$, and $\mathrm{d}(\mathbf{x}, \mathbf{z}) = 2$. On the other hand if $A = \{0, 1, 2, 3, 4, 5, 6\}$, $\mathbf{p} = 24601$ and $\mathbf{q} = 54321$, then $\mathrm{d}(\mathbf{p}, \mathbf{q}) = 3$.

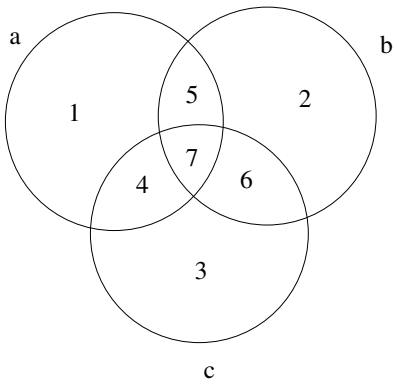
§

Theorem 2. Let $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ be words of length n over A. Then the triangular inequality for their mutual Hamming distance holds, that is

$$\mathrm{d}(\mathbf{x}, \mathbf{z}) \leq \mathrm{d}(\mathbf{x}, \mathbf{y}) + \mathrm{d}(\mathbf{y}, \mathbf{z})$$

Proof. Let $a = \mathrm{d}(\mathbf{x}, \mathbf{z})$, $b = \mathrm{d}(\mathbf{x}, \mathbf{y})$, and $c = \mathrm{d}(\mathbf{y}, \mathbf{z})$. We have $a \geq 0$, $b \geq 0$ and $c \geq 0$. What this theorem states is obvious when $a = 0$. If $a > 0$, then either $b = 0$ or $b > 0$; if the former is the case, that is $b = 0$, then $a = c$ and the theorem is true. If both $a > 0$ and $b > 0$, then either $c = 0$ or $c > 0$; if $c = 0$, then $a = b$ and the theorem is again true. But if $a > 0$, $b > 0$ and $c > 0$, then a, b and c may come from some of the diffences in common, as could be shown in the following Venn diagram.

Figure 2 *Common differences among a, b and c.*



Let (x, y) be the differences in common between distances x and y, and similarly (x, y, z) those among x, y and z. Then from Figure 2 the area 1 is (a); 2, (b); 3, (c); 4, (a, c); 5, (a, b); 6, (b, c); and 7, (a, b, c). Then, $\mathrm{d}(\mathbf{x}, \mathbf{z})$ arises from the differences $(a)+(c)+(a, b)+(b, c)$, $\mathrm{d}(\mathbf{x}, \mathbf{y})$ from $(a)+(b)+(a, c)+(b, c)$, $\mathrm{d}(\mathbf{y}, \mathbf{z})$ from $(b) + (c) + (a, c) + (a, b)$, and therefore $\mathrm{d}(\mathbf{x}, \mathbf{y}) + \mathrm{d}(\mathbf{y}, \mathbf{z})$ gives $(a) + (b) + (c) + (a, c) + (a, b) + (b, c)$, which is never less than in the case of $\mathrm{d}(\mathbf{x}, \mathbf{y})$ and hence the theorem is again true. This exhausts all the cases and the theory is proved. □

Definition 7. The *minimum distance-* or *nearest neighbour* decoding rule decodes $\mathbf{x}$ to $\mathbf{c_x}$ if $\mathrm{d}(\mathbf{x}, \mathbf{c_x}) = \min_{\mathbf{c}\in C} \mathrm{d}(\mathbf{x}, \mathbf{c})$.

§

Exercise 6. A binary code is $C = \{001, 010, 100\}$. If code words are sent over a memoryless binary channel whose channel probabilities are $p_{00} = 0.15$ and $p_{11} = 0.5$, use the maximum likelihood decoding rule to decode the word 111. Then decode 111 again using the nearest neighbour decoding rule.

§

Exercise 7. Our binary code is $C = \{01010, 10101, 10011, 00110\}$. Use the NN (nearest neighbour) decoding rule, decode the words 00000, 11111, 01001, 11011 and 00100.

§

Theorem 3. The maximum likelihood decoding rule and the minimum distance decoding rule is the same for a BSC with cross-over probability $p < \frac{1}{2}$.

Proof. From Theorem 1, when $p < \frac{1}{2}$, gives

$$p^0(1-p)^n > \cdots > p^n(1-p)^0$$

Thus the less the distance the more the likelihood, and thus the theorem is proved. □

Definition 8. Let C be a code containing at least two words. Then, the *minimum distance* or the *distance* of C is

$$\mathrm{d}(C) = \min\{\mathrm{d}(\mathbf{x}, \mathbf{y}) | \mathbf{x}, \mathbf{y} \in C, \mathbf{x} \neq \mathbf{y}\}$$

A code of length n, size m, and distance d is called an (n, m, d)-*code.*

§

Exercise 8. Consider a (n, a, n)-code where $n \geq 2$. Find the value of a.

§

Definition 9. Let a code word be of length n. Then, an *error vector* of weight k is a word containing all the k errors occured taking the value of 1 in their corresponding positions with the remaining positions of the word being zero. An error vector is also called an *error word* or an *error pattern.*

§

Definition 10. An error vector is said to be *detected* by a code if $a + e$ is not a code word for any code word a. If there exists some code word a such that $a + e$ is also a code word, we say that the error vector e goes *undetected.*

§

Definition 11. Let a received word $\mathbf{x}$ differ from the actual code word sent $\mathbf{c}$ by e errors. Then the corresponding code C is said to be u-*error-detecting* if $\mathbf{x}$ is not a code word whenever $1 \leq e \leq u$. Moreover, C is *exactly* u-*error-detecting* if it is u-*error-detecting* but not $(u+1)$-*error-detecting.*

§

Theorem 4. A code C is u-error-detecting if and only if $\mathrm{d}(C) \geq u+1$.

Proof. Let $\mathbf{c} \in C$. If $\mathrm{d}(C) \geq u+1$, then $\mathbf{x}$ such that $1 \leq \mathrm{d}(\mathbf{x}, \mathbf{c}) \leq u < \mathrm{d}(C)$ implies that $\mathbf{x} \notin C$, therefore C is u-error-detecting. On the other hand, if $\mathrm{d}(C) < u+1$, that is $\mathrm{d}(C) \leq u$, then there exist $\mathbf{x}_1, \mathbf{x}_2 \in C$ such that $1 \leq \mathrm{d}(C) \leq \mathrm{d}(\mathbf{x}_1, \mathbf{x}_2) \leq u$, then it is possible to send $\mathbf{c}_1 \in C$ and incur errors such that $1 \leq \mathrm{d}(\mathbf{x}, \mathbf{c}_1) = \mathrm{d}(\mathbf{c}_2, \mathbf{c}_1) \leq u$ and $\mathbf{x} = \mathbf{c}_2$, hence C is not a u-error-detecting code. □

Corollary 4[1]. A code with distance d is exactly $(d-1)$-error-detecting.

§

Definition 12. Let v be a positive integer and assuming the incomplete decoding rule is used. Then a code C is said to be *v-error-correcting* if the minimum distance decoding can correct for it up to v errors. It is said to be *exactly v-error-correcting* if it is v-error-correcting but not $(v+1)$-error-correcting.

§

Theorem 5. A code C is v-error-correcting if and only if $\mathrm{d}(C) \geq 2v+1$.

Proof. Suppose that $\mathrm{d}(C) \geq 2v+1$. Let $\mathbf{c} \in C$ be the code word sent, $\mathbf{x}$ the word received, and e errors occurred such that $e \leq v$. Then $\mathrm{d}(\mathbf{x}, \mathbf{c}) \leq v$, and if C is not to be v-error-correcting there must be some $\mathbf{c}_1, \mathbf{c}_2 \in C$ such that $\mathrm{d}(\mathbf{x}, \mathbf{c}_1) + \mathrm{d}(\mathbf{x}, \mathbf{c}_2) \leq 2v$. But since $\mathrm{d}(C) \geq 2v+1$, which means that $\mathrm{d}(\mathbf{x}, \mathbf{c}_1) + \mathrm{d}(\mathbf{x}, \mathbf{c}_2) \geq 2v+1$ for all $\mathbf{c}_1, \mathbf{c}_2 \in C$, it follows that C must be v-error-correcting.

Next, suppose that C is v-error-correcting and $\mathrm{d}(C) < 2v+1$. Then $\mathrm{d}(C) \leq 2v$, that is to say, there exist $\mathbf{c}_1, \mathbf{c}_2 \in C$ such that $\mathrm{d}(\mathbf{c}_1, \mathbf{c}_2) \leq 2v$. This means that there exist $\mathbf{x}$ such that $\mathrm{d}(\mathbf{x}, \mathbf{c}_1) + \mathrm{d}(\mathbf{x}, \mathbf{c}_2) = \mathrm{d}(\mathbf{c}_1, \mathbf{c}_2) \leq 2v$, hence C is not v-error-correcting. This contradicts what we have supposed earlier, therefore necessarily $\mathrm{d}(C) \geq 2v+1$. □

Corollary 5[1]. A code with distance d is exactly $\left\lfloor \frac{(d-1)}{2} \right\rfloor$-error-correcting code, where $\lfloor x \rfloor$ is the greatest integer less than or equal to x.

§

Bibliography

San Ling and Chaoping Xing. *Coding theory, a first course.* Cambridge University Press, 2004

L R Vermani. *Elements of algebraic coding theory.* Chapman & Hall, 1996

Exercises for Error and distance

14^{th} January, 2007

1. Binary code words from the code $\{000, 010, 101, 110, 111\}$ are sent over a binary symmetric channel (BSC) with cross-over probability $p = 0.02$. Decode using the maximum likelihood decoding rule the following words: 001, 011 and 001.

2. Let a memoryless binary channel have channel probabilities $p_{00} = 0.8$ and $p_{11} = 0.9$, where p_{ij} is the probability that j is received when i is sent. Suppose the code words being sent over this channel are from the binary code $\{000, 100, 110, 011, 111\}$. Decode the words 001, 010 and 101 with the use of the maximum likelihood decoding rule.

3. Consider a binary code $C = \{010, 110, 101\}$.

a. Use the nearest neighbour decoding rule to decode the received word 000.

b. Let our channel be binary and memoryless with the probabilities $p_{00} = 0.2$ and $p_{11} = 0.5$. Decode the received word 000 using the maximum likelihood decoding rule.

4. Decode using the nearest neighbour decoding rule for the binary code

$$C = \{10110, 11000, 10100, 10011, 11011\}$$

the received words 00000, 00011, 01101, 01111 and 10011.

5. Use the nearest neighbour decoding rule to decode for the ternary code

$$C = \{01122, 10021, 20210, 22200\}$$

the received code words 00122, 12001, 20111 and 22000.

6. Construct the incomplete maximum likelihood decoding (IMLD) table for the binary code $C = \{000, 010, 101, 110, 111\}$.

7. Find the number of binary $(n, 2, n)$-codes, $n \geq 2$, where for a (n, m, d)-code n is the length of the word, m the size of the dictionary and d the distance of the code.

8. Consider the binary repetition code of length 6 sent over a binary symmetric channel which has symbol error probability p. Find the word error probability of the code.

9. Consider q-ary $(3, m, 2)$-codes, where $q \geq 2$. Find the range which m may take.

10. Let $A_q(n, d)$ represent the largest value of m such that there exists a q-ary (n, m, d)-code. Find $A_q(n, 1)$ and $A_q(n, n)$. Then find $A_q(3, 2)$ for any integer $q \geq 2$.

11. Find the upper bound of m for the q-ary $(q + 1, m, 3)$-code.

12. Consider a balanced block (b, v, r, k, λ)-design, where b is the number of subsets B_i of a set S of v elements, each point appears in exactly r blocks, each block comprises exactly k points, and each pair of points occurs together in exactly λ blocks. Here B_i are call *blocks*, and S is said to contain v *varieties*. Show that $bk = vr$ and $r(k - 1) = \lambda(v - 1)$

13. A *permutation* of a set S is a one-to-one mapping from S to itself. Two q-ary codes are said to be *equivalent* to each other if one can be obtained from the other by permutation of the positions of the code, or permutation of the symbols appearing in a fixed position, or any combination of both. Show that the binary codes $C_1 = \{00100, 00011, 11000, 11111\}$ and $C_2 = \{00000, 01101, 10110, 11011\}$ are equivalent. Then show that the ternary code $C_3 = \{012, 120, 201\}$ is equivalent to the ternary repetition code of length 3, $C_4 = \{000, 111, 222\}$.

14. Prove that a sphere of radius r in $\mathbf{F}_q^n$, $0 \leq r \leq n$, contains exactly

$$\binom{n}{0} + \binom{n}{1}(q-1) + \cdots + \binom{n}{r}(q-1)^r$$

vectors.

Reference

Raymond Hill. *A first course in coding theory.* Clarendon, 1986

San Ling and Chaoping Xing. *Coding theory, a first course.* Cambridge University Press, 2004

Entropy and mutual information

4^{th} November 2005

Definition 13. A *probability space* is a triple $(\mathrm{S}, \mathrm{B}, \mathrm{P})$ on the domain S, which is a nonempty set called the *sample space*, where (S, B) is a measurable space, B is a Borel field of subsets of S, and P is a measure on S with the property that $\mathrm{P}(\mathrm{S}) = 1$ and, for all disjoint $E_i \in \mathrm{B}$,

$$\mathrm{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathrm{P}(E_i)$$

In other words, P is a nonnegative function defined for all events $E_i \in \mathrm{B}$, and B measurable subsets of S. Further, a *random variable* X is a function mapping S into some set R, called the range of X. For convenience, we shall also use X to represent both the function and its own range, that is X is a function which maps S into X. If S is discrete and f is some real-valued function defined on S, then both X and f(X) are two different random variables, and the expectation of the latter is given by,

$$\mathrm{E}\left[\mathrm{f}(\mathrm{X})\right] = \sum_{x} \mathrm{p}(x)\mathrm{f}(x)$$

§

Definition 14. Let $\mathrm{p}(x)$ be the probability that $x \in \mathrm{X}$ occurs, similarly $\mathrm{p}(y)$ that $y \in \mathrm{Y}$ does-, while $\mathrm{p}(x, y)$ that both $x \in \mathrm{X}$ and $y \in \mathrm{Y}$ do occur. Then,

$$\mathrm{p}(x|y) = \frac{\mathrm{p}(x, y)}{\mathrm{p}(y)} \tag{1}$$

and

$$\mathrm{p}(y|x) = \frac{\mathrm{p}(x, y)}{\mathrm{p}(x)} \tag{2}$$

§

Definition 15. A *Markov chain* is a set of random variable X_t, where $t = 0, 1, \ldots$, such that,

$$\mathrm{P}\left(\mathrm{X}_t = j | \mathrm{X}_0 = i_0, \ldots, \mathrm{X}_{t-1} = i_{t-1}\right) = \mathrm{P}\left(\mathrm{X}_t = j | \mathrm{X}_{t-1} = i_{t-1}\right)$$

In other words, given the present state, the next state is conditionally independent of the past.

§

Definition 16. A subset $\mathrm{K} \subseteq \mathrm{E}^n$, where E^n is the Euclidean space of n dimensions, is called *convex* if the line segment joining any two points in K

is contained in K. Let the two points be x_1 and x_2, then the line segment joining them together is $x = tx_1 + (1-t)x_2$, where $0 \leq t \leq 1$.

§

Definition 17. A point x is said to be a *convex combination* of points $x_1, \ldots, x_m$ if there exist nonnegative scalars $\alpha_1, \ldots, \alpha_m$ such that $\sum \alpha_i = 1$ and $\sum \alpha_i x_i = x$. The set of all convex combinations of x_i, $i = 1, \ldots, m$, is called the *convex hull* of $\{x_i\}$.

§

Definition 18. Let f be a real-valued function, and let K be a convex subset of the domain of f. Then f is said to be *convex cup* if, for every $x_1, x_2 \in$ K and $0 \leq t \leq 1$,

$$\mathrm{f}(tx_1 + (1-t)x_2) \leq t\mathrm{f}(x_1) + (1-t)\mathrm{f}(x_2) \tag{3}$$

It is said to be *strictly convex cup* if strict inequality holds in Equation 3 whenever $x + 1 \neq x_2$. Similarly, f is said to be *convex cap* if,

$$\mathrm{f}(tx_1 + (1-t)x_2) \geq t\mathrm{f}(x_1) + (1-t)\mathrm{f}(x_2) \tag{4}$$

that is to say, if $-f$ is convex cup. It is said to be *strictly convex cap* if strict inequality holds in Equation 4 whenever $x_1 \neq x_2$. Convex cap is also known as *concave*. Geometrically speaking, f is convex cup if and only if all its chords lie above or on the graph of f, and f is concave if and only if all its chords lie below or on the graph of the same.

§

Definition 19. Let K be some interval in E^1, and let $\mathrm{F}(x)$ be a probability distribution concentrated on K such that $\mathrm{P}(\mathrm{X} \leq x) = \mathrm{F}(x)$. Then, if the expectation E(X) exists, and if $\mathrm{f}(x)$ is a convex cup function, then,

$$\mathrm{E}(\mathrm{f}(\mathrm{X})) \geq \mathrm{f}(\mathrm{E}(\mathrm{X})) \tag{5}$$

If f is strictly convex cup, then strict inequality holds in Equation 5. Similarly, if f is convex cap, then,

$$\mathrm{E}(\mathrm{f}(\mathrm{X})) \leq \mathrm{f}(\mathrm{E}(\mathrm{X}) \tag{6}$$

If f is strictly convex cap, then strict inequality holds in Equation 6.

§

Example 7. Suppose that in Definition 19 there is a mass distribution placed on the graph of f, then Equation 5 says that the overall centre of mass will lie above or on the graph, while Equation 6 says that it will lie below it.

Entropy is a measure of uncertainty of many events as a single value. We derive it from Axiom's 1 and 2.

Axiom 1. If the events are all equally likely, then the uncertainty function $\mathrm{H}\left(\frac{1}{m}, \ldots, \frac{1}{m}\right)$ is monotonously increasing with m.

§

Axiom 2. If $\{E_1^1, \ldots, E_m^1\}$ and $\{E_1^2, \ldots, E_n^2\}$ are statistically independent sets of equally likely disjoint events, then the uncertainty of the sets of events $\{E_i \cap E_j; i = 1, \ldots, m; j = 1, \ldots, n\}$ is

$$\mathrm{H}\left(\frac{1}{mn}, \ldots, \frac{1}{mn}\right) = \mathrm{H}\left(\frac{1}{m}, \ldots, \frac{1}{m}\right) + \mathrm{H}\left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$$

That is to say, $\mathrm{h}(mn) = \mathrm{h}(m) + \mathrm{h}(n)$, where $\mathrm{h}(m) = \mathrm{H}\left(\frac{1}{m}, \ldots, \frac{1}{m}\right)$.

§

Definition 20. Let the set of m possible disjoint events be

$$\mathrm{E} = \{E_1, \ldots, E_m\}$$

We call an *apriori probability* of E_i, $\mathrm{p}(E_i)$, where $1 \leq i \leq m$ and $\sum_{i=1}^{m} \mathrm{p}(E_i) = 1$. The *uncertainty function* or the *entropy function*, $\mathrm{H}(\mathrm{p}(1), \ldots, \mathrm{p}(m))$ obeys Axiom's 1 and 2.

§

The entropy of a random variable x gives a measure of the amount of *information* obtained from an observation of x. It also represents the *randomness* of x and our *uncertainty* about x. The less probable an event is, the more information we receive when it occurs.

Theorem 6. The entropy of a set of m equally likely events is $\mathrm{h}(m) = \lambda \log_c m$, where λ is a positive constant and $c > 1$.

Proof. Proving Theorem 6 amounts to proving that Axiom's 1 and 2 are satisfied if and only if $\mathrm{h}(m) = \lambda \log_c m$. The two axioms say that $\mathrm{h}(m)$ is monotonously increasing in m and

$$\mathrm{h}(mn) = \mathrm{h}(m) + \mathrm{h}(n) \tag{7}$$

According to Equation 7, if $m = n = 1$, then $\mathrm{h}(1) = \mathrm{h}(1) + \mathrm{h}(1)$, which implies that $\mathrm{h}(1) = 0$. From this together with both axioms above, $\mathrm{h}(m) = \lambda \log_c m$ is sufficient as a solution.

Next, we must prove that this solution is necessarily the only solution. Let a, b and c be positive integers, and $a, b, c > 1$. Then there exists a unique integer d such that

$$c^d \leq a^b < c^{d+1} \tag{8}$$

From Equation 8 it follows that,

$$d \log c \leq b \log a < (d+1) \log c$$

and therefore,

$$\frac{d}{b} \leq \frac{\log a}{\log c} < \frac{d+1}{b} \tag{9}$$

Since $\mathrm{h}(m)$ is monotonously increasing, from Equation 8 we have,

$$\mathrm{h}(c^d) \leq \mathrm{h}(a^b) < \mathrm{h}(c^{d+1})$$

Then from Equation 7, $d\mathrm{h}(c) \leq b\mathrm{h}(a) < (d+1)\mathrm{h}(c)$. And since $\mathrm{h}(m)$ is monotonously increasing,

$$\frac{d}{b} \leq \frac{\mathrm{h}(a)}{\mathrm{h}(c)} < \frac{d+1}{b} \tag{10}$$

From Equation's 9 and 10 it follows that,

$$\left| \frac{\log a}{\log c} - \frac{\mathrm{h}(a)}{\mathrm{h}(c)} \right| < \frac{1}{b}$$

And, since b is arbitrary positive integer,

$$\frac{\mathrm{h}(a)}{\mathrm{h}(c)} = \frac{\log a}{\log c}$$
$$\frac{\mathrm{h}(a)}{\log a} = \frac{\mathrm{h}(c)}{\log c}$$

Since a and c are arbitrary,

$$\frac{\mathrm{h}(a)}{\log a} = \lambda = \frac{\mathrm{h}(c)}{\log c}$$

Therefore, necessarily $\mathrm{h}(m) = \lambda \log_c m$ is the only solution. ¶

Axiom 3. The total uncertainty of events does not depend on the method of indication.

§

Axiom 4. The uncertainty measure is a continuous function with regard to the probabilities within it.

§

Example 8. Let a set E of m disjoint events be

$$\{E_1, \ldots, E_m\}$$

Let j_i, $i = 0, \ldots, n$, be integers and $0 = j_0 \leq j_1 < j_2 \cdots < j_n = m$, and E be divided into n sets of events, namely,

$$\mathrm{G}_1 = \{E_1, \ldots, E_{j_1}\}$$
$$\mathrm{G}_2 = \{E_{j_1+1}, \ldots, E_{j_2}\}$$
$$\vdots$$
$$\mathrm{G}_n = \{E_{j_{n-1}+1}, \ldots, E_m\}$$

If we indicate firstly the group, and then the event within that group, then the uncertainty becomes,

$$\mathrm{H}(\mathrm{p}(E_1),\ldots,\mathrm{p}(E_m)) = \mathrm{H}(\mathrm{p}(\mathrm{G}_1),\ldots,\mathrm{p}(\mathrm{G}_n)) + \sum_{i=1}^{n} \mathrm{p}(\mathrm{G}_i)\mathrm{H}(\mathrm{p}(E_{j_{i-1}+1}|\mathrm{G}_i),\ldots,\mathrm{p}(\mathrm{E}_{j_i}|\mathrm{G}_i)) \quad (11)$$

The grouping axiom, Axiom 3, lets us express the uncertainty when all the event probabilities are rational. By grouping equally likely events together and then consider each of the groups as a single event, it gives us the ability to deal with events which are not equally likely. Example 9 gives an example how this is done. Then Axiom 4 extends Axiom 3 to cover also irrational probabilities, and Equation 12 is the result.

Example 9. As in Example 8, let a set of disjoint events be

$$\mathrm{E} = \{E_1,\ldots,E_m\}$$

and let $\mathrm{p}(E_i) = \frac{1}{m}$, $i = 1,\ldots,m$. Also, let the groups of events $\mathrm{G}_1,\ldots,\mathrm{G}_n$ be defined the same way therein. Let n_k be the number of events in G_k. Then $n_k = j_k - j_{k-1}$ and $\mathrm{p}(\mathrm{G}_k) = \frac{n_k}{m}$, for $k = 1,\ldots,n$, and also $\mathrm{p}(E_i|\mathrm{G}_k) = \frac{1}{n_k}$, for $j_{k-1} < i \leq j_k$. Then Equation 11 yields,

$$\mathrm{h}(m) = \mathrm{H}(\mathrm{p}(\mathrm{G}_1),\ldots,\mathrm{p}(\mathrm{G}_n)) + \sum_{i=1}^{n} \mathrm{p}(\mathrm{G}_i)\mathrm{h}(n_i)$$

And since from Theorem 6, $\mathrm{h}(m) = \lambda \log_c m$, we have,

$$\begin{aligned}\mathrm{H}(\mathrm{p}(\mathrm{G}_1),\ldots,\mathrm{p}(\mathrm{G}_n)) &= -\sum_{i=1}^{n} \mathrm{p}(\mathrm{G}_i)(\mathrm{h}(n_i) - \mathrm{h}(m)) \\ &= -\sum_{i=1}^{n} \mathrm{p}(\mathrm{G}_i)\left(\lambda \log \frac{n_i}{m}\right) \\ &= -\lambda\left(\sum_{i=1}^{n} \mathrm{p}(\mathrm{G}_i)\log \mathrm{p}(\mathrm{G}_i)\right) \quad (12)\end{aligned}$$

Example 10. From $\mathrm{h}(m) = \lambda \log_c m$, if we let $\lambda = \log_b c$, then $\mathrm{h}(m) = \log_b m$. In other words, the scale factor λ can be absorbed in the base of the logarithm.

Theorem 7. Let $\{p_1,\ldots,p_m\}$ be a set of probabilities such that $\sum_{i=1}^{m} p_i = 1$. Then, †

$$\mathrm{H}(p_1,\ldots,p_m) = -\sum_{i=1}^{m} p_i \log p_i \quad (13)$$

Proof. This is the results from Example's 8 and 9, and the scale factor λ disappears in a manner similar to that shown by Example 10. ¶

† Some times the entropy function is defined instead by $\mathrm{H}(p_1,\ldots,p_m) = \sum_{i=1}^{m} p_i \log \frac{1}{p_i}$, but this is obviously the same as our Equation 13 since $\log x^{-1} = -\log x$.

Example 11. If the base of the logarithm in Equation 13 is 2, the unit of the entropy is *bit*. On the other hand if this base is e, that is to say, if we use natural logarithms, then the uncertainty has the unit of *nat*. From this, one may see that one nat is equal to $\log_2 e$ bits, which is approximately 1.443 bits. The term *bit* comes from *binary digit*, the term *nat* from *natural digit*.

Definition 21 explains what is meant by *conditional entropy*. Starting from Equation 14, which is an equation for conditional entropy when y is given, we obtain the overall conditional entropy in Theorem 8. For any pair of sets X and Y given, H(X|Y) gives the amount of uncertainty remaining about X after Y has been observed.

Definition 21. The *conditional entropy* of X, given some $y \in$ Y, is,

$$\mathrm{H}(\mathrm{X}|y) = -\sum_{x} \mathrm{p}(x|y) \log \mathrm{p}(x|y) \tag{14}$$

Then the conditional entropy H(X|Y) is the expectation, or average value, of H(X|y) over the range Y. In other words,

$$\mathrm{H}(\mathrm{X}|\mathrm{Y}) = \sum_{y} \mathrm{p}(y)\mathrm{H}(\mathrm{X}|y) \tag{15}$$

§

Theorem 8. The conditional entropy is,

$$\mathrm{H}(\mathrm{X}|\mathrm{Y}) = -\sum_{x,y} \mathrm{p}(x,y) \log \mathrm{p}(x|y)$$

Proof. Putting the equation of conditional entropy when y is given, Equation 14, into the overall conditional entropy equation, Equation 15, we get,

$$\begin{aligned} \mathrm{H}(\mathrm{X}|\mathrm{Y}) &= \sum_{y} \mathrm{p}(y)\mathrm{H}(\mathrm{X}|y) \\ &= -\sum_{y} \mathrm{p}(y) \sum_{x} \mathrm{p}(x|y) \log \mathrm{p}(x|y) \end{aligned}$$

Then from Equation 1 of Definition 14, $\mathrm{p}(y)\mathrm{p}(x|y) = \mathrm{p}(x,y)$, and so,

$$\mathrm{H}(\mathrm{X}|\mathrm{Y}) = -\sum_{x,y} \mathrm{p}(x,y) \log \mathrm{p}(x|y)$$

¶

Theorem 9. Let X, Y and Z be discrete random variables. For each $z \in \mathrm{Z}$, let $\mathrm{E}(z) = \sum_{x,y} \mathrm{p}(y)\mathrm{p}(z|x,y)$. Then,

$$\mathrm{H(X|Y)} \leq \mathrm{H(Z)} + \mathrm{E}(\log \mathrm{E})$$

Proof.

$$\begin{aligned}\mathrm{H(X|Y)} &= -\mathrm{E}\left[\log \mathrm{p}(x|y)\right] \\ &= -\sum_{x,y,z} \mathrm{p}(x,y,z) \log \mathrm{p}(x|y) \\ &= -\sum_{z} \mathrm{p}(z) \sum_{x,y} \frac{\mathrm{p}(x,y,z)}{\mathrm{p}(z)} \log \mathrm{p}(x|y)\end{aligned}$$

Because

$$\frac{\mathrm{p}(x,y,z)}{\mathrm{p}(z)} = \mathrm{p}(x,y|z)$$

is a probability distribution, that is a convex cap function, we may apply Equation 6, namely Jensen's inequality for convex cap, from Definition 19. Hence,

$$\begin{aligned}\mathrm{H(X|Y)} &\leq \sum_{z} \mathrm{p}(z) \log \left(\frac{1}{\mathrm{p}(z)} \sum_{x,y} \frac{\mathrm{p}(x,y,z)}{\mathrm{p}(x|y)} \right) \\ &= \sum_{z} \mathrm{p}(z) \log \frac{1}{\mathrm{p}(z)} + \sum_{z} \mathrm{p}(z) \log \sum_{x,y} \frac{\mathrm{p}(x,y,z)}{\mathrm{p}(x|y)}\end{aligned}$$

But,

$$\frac{\mathrm{p}(x,y,z)}{\mathrm{p}(x|y)} = \frac{\mathrm{p}(x,y,z)\mathrm{p}(y)}{\mathrm{p}(x,y)} = \mathrm{p}(y)\mathrm{p}(z|x,y)$$

hence the statement above is proved. ¶

Corollary 9[1]. Let X and Y be random variables each of which takes values in the set $\{x_1, \ldots, x_r\}$. Let $\mathrm{P}_e = \mathrm{P}(\mathrm{X} \neq \mathrm{Y})$. Then,

$$\mathrm{H(X|Y)} \leq \mathrm{H}(\mathrm{P}_e) + \mathrm{P}_e \log(r-1)$$

Proof. From Theorem 9, let $\mathrm{Z} = 0$ if $\mathrm{X} = \mathrm{Y}$, and let $\mathrm{Z} = 1$ if $\mathrm{X} \neq \mathrm{Y}$. Then $\mathrm{E}(0) = 1$ and $\mathrm{E}(1) = r - 1$. ¶

Theorem 10. The maximum uncertainty occurs when the events are equiprobable.

Proof. Since,

$$\begin{aligned}\mathrm{H}\left(\frac{1}{m}, \ldots, \frac{1}{m}\right) - \mathrm{H}(p_1, \ldots, p_m) &= \log_b m + \sum_{i=1}^{m} p_i \log_b p_i \\ &= \log_b e \sum_{i=1}^{m} p_i \ln m p_i \\ &\geq \log_b e \sum_{i=1}^{m} p_i \left(1 - \frac{1}{m p_i}\right) = 0\end{aligned}$$

it being the case that $\ln \frac{1}{x} \geq 1 - x$. Therefore $\mathrm{H}(p_1, \ldots, p_m)$ is maximised when $p_i = \frac{1}{m}$, for all $i = 1, \ldots, m$. ¶

Example 12. Figure 3 shows that $\ln x \leq x - 1$, while Figure 4 shows that such inequality does not exist when the logarithm in question is of base 10.

Figure 3 *Plots of* $\ln x$ *and* $x - 1$*, which show that* $\ln x \leq x - 1$.

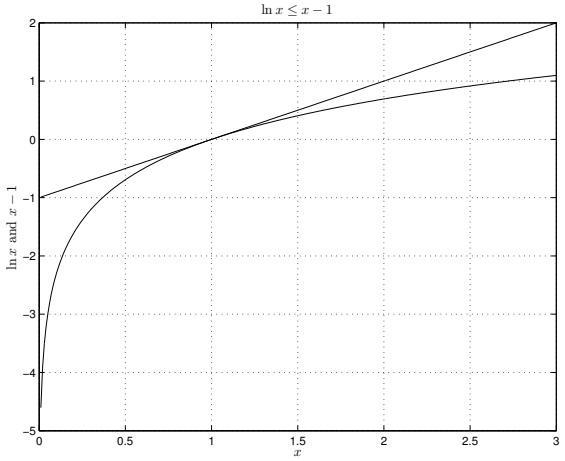
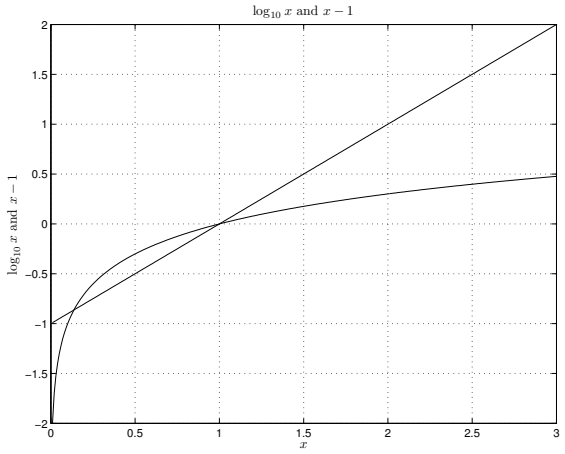


Figure 4 *Graphs of* $y = \log x$ *and* $y = x - 1$*, which show that the latter is no bound for the values of the former.*



Example 13. Figure 5 confirms for us how $\ln \frac{1}{x} \geq 1 - x$, whereas Figure 6 tells us that this is the case for $\log \frac{1}{x}$.

Figure 5 *Plots showing* $\ln \frac{1}{x}$ *and* $1 - x$*, which show that* $\ln \frac{1}{x} \geq 1 - x$.

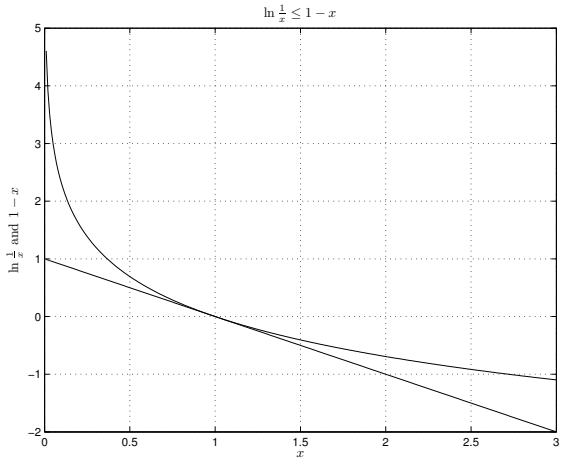
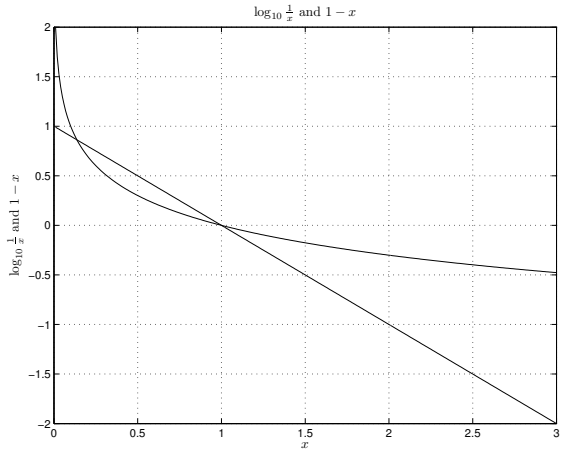


Figure 6 *Graphs showing* $y = \log \frac{1}{x}$ *and* $y = 1 - x$*, from which it is clear the latter gives no bounds for the former.*

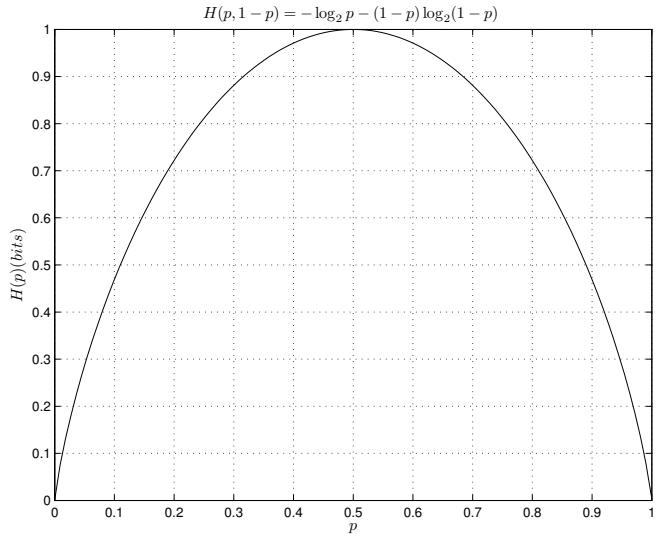


Example 14. Consider two events with probabilities p and $1 - p$. The entropy function is then,

$$\mathrm{H}(p, 1-p) = -p \log p - (1-p) \log(1-p)$$

Whenever the occurrence of either event become certainty, the entropy function would become zero. Mathematically we see that $\lim_{p \to 0} p \log p = 0$ and $\lim_{p \to 1} p \log p = 0$. Figure 7 shows a plot of the values of the entropy function for two events. Base-2 logarithm is used here.

Figure 7 *The entropy function of two events with probabilities* p *and* $1 - p$*.*



Definition 22. The *mutual information* is $\mathrm{I(X;Y)} = \mathrm{H(X)} - \mathrm{H(X|Y)}$. It represents the information provided about X by Y.

§

Example 15. Alternatively, the mutual information may take the following form, *cf* Definition 14,

$$\begin{aligned}\mathrm{I(X;Y)} &= \sum_{x,y} \mathrm{p}(x,y) \log \frac{\mathrm{p}(x|y)}{\mathrm{p}(x)} \\ &= \sum_{x,y} \mathrm{p}(x,y) \log \frac{\mathrm{p}(x,y)}{\mathrm{p}(x)\mathrm{p}(y)} \\ &= \sum_{x,y} \mathrm{p}(x,y) \log \frac{\mathrm{p}(y|x)}{\mathrm{p}(y)}\end{aligned}$$

That is to say, $\mathrm{I(X;Y)}$ is the average taken over the X, Y sample space of the random variable $\mathrm{I}(x;y)$ such that,

$$\mathrm{I}(x;y) = \log \frac{\mathrm{p}(x|y)}{\mathrm{p}(x)} = \log \frac{\mathrm{p}(x,y)}{\mathrm{p}(x)\mathrm{p}(y)} = \log \frac{\mathrm{p}(y|x)}{\mathrm{p}(y)}$$

Theorem 11. For any discrete random variables X and Y, $\mathrm{I(X;Y)} \geq 0$. Moreover, $\mathrm{I(X;Y)} = 0$ if and only if X and Y are independent.

Proof. From one of our formulae for the mutual information and from

Jensen's inequality,

$$\begin{aligned} \mathrm{I(X;Y)} &= -\sum_{x,y} \log \frac{\mathrm{p}(x)\mathrm{p}(y)}{\mathrm{p}(x,y)} \\ &\geq \log \sum_{x,y} \mathrm{p}(x)\mathrm{p}(y) = \log 1 = 0 \end{aligned}$$

Furthermore, the equality sign holds if and only if $\mathrm{p}(x)\mathrm{p}(y) = \mathrm{p}(x,y)$ for all x and y, that is to say, when X and Y are independent of each other. ¶

Example 16. From our formulae of the mutual information, we may see that,

$$\mathrm{I(X;Y)} = \mathrm{I(Y;X)}$$

and

$$\mathrm{I(X;Y)} = \mathrm{H(Y)} - \mathrm{H(Y|X)}$$

Also,

$$\mathrm{I(X;Y)} = \sum_{x,y} \mathrm{p}(x,y) \log \frac{1}{\mathrm{p}(x,y)}$$

Definition 23. Let X, Y and Z be three random variables. Then the mutual information $\mathrm{I(X,Y;Z)}$ is given by,

$$\mathrm{I(X,Y;Z)} = \mathrm{E}\left(\log \frac{\mathrm{p}(z|x,y)}{\mathrm{p}(z)}\right) = \sum_{x,y,z} \mathrm{p}(x,y,z) \log \frac{\mathrm{p}(z|x,y)}{\mathrm{p}(z)}$$

This mutual information is the amount of information X and Y provide about Z.

§

Theorem 12. Let X, Y and Z be three random variables. Then we have $\mathrm{I(X,Y;Z)} \geq \mathrm{I(Y;Z)}$, where the equality holds if and only if $\mathrm{p}(z|x,y) = \mathrm{p}(z|y)$ for all (x,y,z) such that $\mathrm{p}(x,y,z) > 0$.

Proof.

$$\begin{aligned} \mathrm{I(Y;Z)} - \mathrm{I(X,Y;Z)} &= \mathrm{E}\left(\log \frac{\mathrm{p}(z|y)}{\mathrm{p}(z)} - \log \frac{\mathrm{p}(z|x,y)}{\mathrm{p}(z)}\right) \\ &= \mathrm{E}\left(\log \frac{\mathrm{p}(z|y)}{\mathrm{p}(z|x,y)}\right) \\ &= \sum_{x,y,z} \mathrm{p}(x,y,z) \log \frac{\mathrm{p}(z|y)}{\mathrm{p}(z|x,y)} \end{aligned}$$

Then using Jensen's inequality, we have,

$$\begin{aligned} \mathrm{I(Y;Z)} - \mathrm{I(X,Y;Z)} &\leq \log \sum_{x,y,z} \mathrm{p}(x,y,z) \frac{\mathrm{p}(z|y)}{\mathrm{p}(z|x,y)} \\ &= \log \sum_{x,y,z} \mathrm{p}(x,y)\mathrm{p}(z|y) = \log 1 = 0 \end{aligned}$$

¶

Theorem 13. Let (X, Y, Z) be a Markov chain. Then,

$$I(X;Z) \leq \begin{cases} I(X;Y) \\ I(Y;Z) \end{cases}$$

Proof. From Theorem 12, $I(X;Z) \leq I(X,Y;Z)$. Because (X,Y,Z) is a Markov chain, $I(X,Y;Z) = I(Y;Z)$. Therefore $I(X;Z) \leq I(Y;Z)$. Next, since (X,Y,Z) is a Markov chain, (Z,Y,X) is also a Markov chain. Hence $I(X;Z) \leq I(X;Y)$. ¶

Bibliography

Solomon W Golomb, Robert E Peile and Robert A Scholtz. *Basic concepts in information theory and coding, The adventures of Secret Agent 00111.* Plenum Press, New York, 1994

Robert J McEliece *The theory of information and coding.* Addison-Wesley, 1977

Group, field and finite field

11^{th} November 2005

Definition 24. A *group* is a non-empty set G together with an operation, called *multiplication*, which associates with each ordered pair x, y of elements in G a third element, their *product*, in G such that,

1. multiplication is *associative*;
2. there exists an *identity element* e in G; and
3. for each element x in G there exists an *inverse* of x.

In other words, for x and y in G there exists xy in G such that,

1. for any x, y and z in G, $x(yz) = (xy)z$;
2. there exists e in G such that $xe = ex = x$; and
3. to each x in G there corresponds x^{-1} in G such that $xx^{-1} = x^{-1}x = e$.

A group is called *Abelian* or *commutative group* if $xy = yx$ for all elements x and y in G. The group G is called a *finite group* if it consists of a finite number of elements, otherwise it is called an *infinite group*. This number of elements of G is called its *order*.

§

Theorem 14. Both the identity e and the inverse x^{-1} of a group G are unique.

Proof. Suppose e^90 is another element in G such that $xe^90 = e^90x = x$ for every x in G, then $e^90 = e^90e = e$, hence the identity element is unique. Suppose for every x in G, that x^90 be another element in G such that $xx^90 = x^90x = e$, then,

$$x^90 = x^90e = x^90(xx^{-1}) = (x^90x)x^{-1} = ex^{-1} = x^{-1}$$

hence the inverse element of G is unique. ¶

Definition 25. A *ring* is an additive Abelian group R which is closed under a second operation, called *multiplication*, in such a manner that,

1. multiplication is *associative*; and
2. multiplication is *distributive*.

That is to say, if x, y and z are any three elements in R, then,

1. $x(yz) = (xy)z$; and
2. $x(y + z) = xy + xz$ and $(x + y)z = xz + yz$.

A ring is called a *commutative ring* if $xy = yx$ for all elements x and y in R. If a ring R has a non-zero element 1 with such a property that $x1 = 1x = x$ for every x, then 1 is called an *identity element*, and R is said to be a *ring with identity*.

§

Definition 26. Let x be an element of R, a ring with identity. Then x is said to be *regular* if its inverse x^{-1} exists, otherwise it is said to be *singular*. Regular elements are also called *invertible-* or *non-singular* elements. Furthermore, R is called a *division ring* if all its non-zero elements are regular.

§

Definition 27. A *field* is a commutative division ring.

§

Example 17. A field, then, is a non-empty set F together with two operations on its elements, namely addition and multiplication, such that for all a, b and c in F, under addition, F is closed, commutative, associative, has a unique identity, has for each of its elements a unique inverse; and under multiplication, F is closed, commutative, associative, has a unique identity, has for each of its elements a unique inverse. Furturemore, F is also distributive.

These properties of field are inherited from the latter's progenitors, since the field is defined by the division ring which itself is defined by the ring which itself is defined by the group. Table 1 shows the sources from which each of the properties of the field is defined.

operator	*property*	*defining definition*
addition	closed	group
	commutative	Abelian group
	associative	group
	identity	group
	inverse	group
multiplication	closed	ring
	commutative	commutative ring
	associative	ring
	identity	ring with identity
	inverse	division ring

Table 1 *The various sources at the places of which the various properties of the field are defined.*

Theorem 15. Consider any two elements a and b in a field F, we have $(-1) \cdot a = -a$.

Proof. Since,

$$(-1) \cdot a + a = (-1) \cdot a + a \cdot 1 = ((-1) + 1) \cdot a = 0 \cdot a = 0$$

and since $a + (-a) = 0$, therefore $(-1) \cdot a = -a$. ¶

Theorem 16. Let a and b be any two elements in a field F. Then $ab = 0$ implies $a = 0$ or $b = 0$.

Proof. If $a \neq 0$, then,

$$0 = a^{-1} \cdot 0 = a^{-1}(ab) = (a^{-1}a)b = 1 \cdot b = b \cdot 1 = b$$

And since a and b are arbitrary, and since $ab = ba$, our statement above is proved. ¶

Definition 28. Let a, b and m be integers, and let $m > 1$. Then a is said to be *congruent to b modulo m*, in other words,

$$a \equiv b (\mathrm{mod}\, m)$$

if $m|(a - b)$, that is to say, m divides $a - b$. The number m is called the *modulus*, and b is called the *residue* of $a (\mathrm{mod}\, m)$. Sometimes b is also called the *principal remainder* of a divided by m, and denoted by $(a (\mathrm{mod}\, m))$. A residue is said to be *common* if $0 \leq b < m$.

§

Theorem 17. Any integer a is congruent to exactly one of $0, 1, \ldots, m - 1$ modulo m.

Proof. Let a and m be integers, and let $m > 1$. Then there exists a unique k such that $a = mk + b$, where $0 \leq b \leq m - 1$. Therefore b is uniquely determined by m and a.
To prove that b is unique, suppose there exist $a = mk_1 + b_1$ and $a = mk_2 + b_2$, where $0 \leq b_1 \leq m - 1$ and $0 \leq b_2 \leq m - 1$, such that $b_1 \neq b_2$. Then, $a - mk_1 \neq a - m_k 2$, and since $m > 1$, therefore $k_1 \neq k_2$. Since k_1 and k_2 are arbitrary, let $k_1 > k_2$ and let $k_1 = k_2 + n$. Then,

$$mk_2 + b_2 = a = m(k_2 + n) + b_1 = mk_2 + b_1 + mn$$

and since $b_1 \geq 0$, $m \geq 0$ and $n > 0$, we have $b_2 \geq m$, which contradicts what we have said earlier, that is $b_2 \leq m - 1$. So, necessarily $b_1 = b_2$. ¶

Theorem 18. Let a, b and m are integers, and let $m > 1$. Then the following properties hold for congruence.

a. $a \equiv b (\mathrm{mod}\, 0)$ implies $a = b$
b. either $a \equiv b (\mathrm{mod}\, m)$ or $a \not\equiv b (\mathrm{mod}\, m)$
c. $a \equiv a (\mathrm{mod}\, m)$
d. $a \equiv b (\mathrm{mod}\, m)$ implies $b \equiv a (\mathrm{mod}\, m)$
e. if $a \equiv b (\mathrm{mod}\, m)$ and $b \equiv c (\mathrm{mod}\, m)$, then $a \equiv c (\mathrm{mod}\, m)$

Let $a \equiv b (\mathrm{mod}\, m)$ and $c \equiv d (\mathrm{mod}\, m)$. Then,

f. $a + c \equiv b + d (\mathrm{mod}\, m)$
g. $a - c \equiv b - d (\mathrm{mod}\, m)$
h. $ac \equiv bd (\mathrm{mod}\, m)$

Further, let k and n be integers. Then,

i. if $a \equiv b (\mathrm{mod}\, m)$, then $ka \equiv kb (\mathrm{mod}\, m)$
j. if $a \equiv b (\mathrm{mod}\, m)$, then $a^n \equiv b^n (\mathrm{mod}\, m)$
k. if $a \equiv b (\mathrm{mod}\, m_1)$ and $a \equiv b (\mathrm{mod}\, m_2)$, then,

$$a \equiv b \, (\mathrm{mod}\, \mathrm{lcm}(m_1, m_2))$$

where $\mathrm{lcm}(x,y)$ is the least common multiple of x and y, that is the smallest z such that there exist positive integers p and q by which $px = qy = z$

l. if $a^k \equiv b^k (\mathrm{mod}\, m)$, then,

$$a \equiv b \left(\mathrm{mod}\, \frac{m}{\gcd(k,m)} \right)$$

From above properties, it follows that,

m. if $a \equiv b (\mathrm{mod}\, m)$, then $P(a) \equiv P(b) (\mathrm{mod}\, m)$, where $P(x)$ is a polynomial.

Properties (a) is called *equivalence*, (b) *determination*, (c) *reflexive*, (d) *symmetry*, and (e) *transition*.

§

Definition 29. We denote by $\mathbf{Z}_m$ or $\mathbf{Z}/(m)$ the set $\{0, \ldots, m-1\}$, where $m > 1$, and define the addition and multiplication on it as,

$$a \oplus b = (a + b (\mathrm{mod}\, m))$$

and

$$a \odot b = (ab (\mathrm{mod}\, m))$$

respectively, and these may be denoted as $a + b$ and respectively ab for simplicity.

§

Example 18. The set $\mathbf{Z}$ together with addition and multiplication introduced in Definition 29 form a ring.

Theorem 19. The ring $\mathbf{Z}_m$ is a field if and only if m is prime.

Proof. First we prove that m being prime implies that $\mathbf{Z}_m$ is a field. Let m be a prime. Then any $a \neq 0$ in $\mathbf{Z}_m$, in other words $0 < a < m$, is prime relative to m. Therefore, there exist two integers u and v, where $0 \leq u \leq m-1$, such that $ua + vm = 1$, which means that $ua \equiv 1 (\mathrm{mod}\, m)$. Hence $u = a^{-1}$, and since this applies for every a in $\mathbf{Z}_m$, it follows that $\mathbf{Z}_m$ is a field.
Next we will prove that if m is not a prime, then $\mathbf{Z}_m$ is no field. Suppose that m is not a prime. Then $m = ab$ for some a and b, where $1 < a < m$ and $1 < b < m$. But $ab = 0$ is in $\mathbf{Z}_m$, and therefore $a = 0$ and $b = 0$. This contradicts the values of a and b given above, thus $\mathbf{Z}_m$ is no field. ¶

Definition 30. We denote by na the element $\sum_{i=1}^{n} a$ for any element a in a ring R and an integer $n \geq 1$.

§

Definition 31. Let F be a field. Then the *characteristic* of F is the least positive integer p such that $p \cdot 1 = 0$, where 1 is the multiplicative identity of F. Where no such p exists, this characteristic is defined to be zero.

By F^* we mean $F \backslash \{0\}$.

§

Theorem 20. The characteristics of a field is either zero or a prime number.

Proof. Consider a field F. Since $1 \cdot 1 = 1 \neq 0$, therefore 1 is not the characteristic of F. Let the characteristic be $p = mn$, where $1 < n < p$ and $1 < m < p$. If $a = m \cdot 1$ and $b = n \cdot 1$, then,

$$a \cdot b = (m \cdot 1)(n \cdot 1) = \left(\sum_{i=1}^{m} 1 \right) \left(\sum_{j=1}^{n} 1 \right) = mn \cdot 1 = p \cdot 1 = 0$$

This implies $a = 0$ and $b = 0$, which contradicts what we had assumed when we started. ¶

Definition 32. Let E and F be two fields, and let F be a subset of E. Then F is called a *subfield* of E if the addition and multiplication of E, when restricted to F, are the same as those of F.

§

Theorem 21. A finite field F of characteristic p contains p^n elements for some integer $n \geq 1$.

Proof. Choose an element α_1 from F^*. Then $0 \cdot \alpha_1, \ldots, (p-1) \cdot \alpha_1$ are pairwisely distinct from one another, for if $i \cdot \alpha_1 = j \cdot \alpha_1$ for some $0 \leq i \leq j \leq p-1$, then $(j-i) \cdot \alpha_1 = 0$. Since p is the characteristic of F, by Theorem 20 p can be either zero or prime. And since $0 \leq j - i \leq p-1$, therefore $j - i = 0$, that is $i = j$.
Next, if $F \backslash \{0 \cdot \alpha_1, \ldots, (p-1) \cdot \alpha_1\}$ is not empty we choose from it α_2. Then $a_1\alpha_1 + a_2\alpha_2$ are pairwise distinct for all $0 \leq a_1, a_2 \leq p-1$, for if $a_1\alpha_1 + a_2\alpha_2$ for some $0 \leq a_1, a_2, b_1, b_2 \leq p-1$, then necessarily $a_2 = b_2$ because otherwise,

$$\alpha_2 = \frac{a_1 - b_1}{b_2 - a_2} \alpha_1$$

which contradicts the way we have chosen α_2. Then it follows that $(a_1, a_2) = (b_1, b_2)$.
Since F is finite, we may continue in this fashion to α_3, α_4, and so on until α_n for some integer n, and find α_j, for all $2 \leq j \leq n$, from $F \backslash \{\sum_{i=1}^{j-1} a_i\alpha_i\}$, where a_i, $i = 1, \ldots, j-1$, are in $\mathbf{Z}_p$.
In the end, $F = \{\sum_{i=1}^{n} a_i\alpha_i\}$, where $a_1, \ldots, a_n$ are in $\mathbf{Z}_p$. In the same manner as above, we may show that $a_1\alpha_1 + \ldots + a_n\alpha_n$ are pairwisely distinct from each other for all a_i in $\mathbf{Z}_p$, where $i = 1, \ldots, n$. Therefore $|F| = p^n$. ¶

Definition 33. Let F be a field. Then the set,

$$F[x] = \left\{ \sum_{i=0}^{n} a_i x^i \right\}$$

where a_i is an element in F and $n \geq 0$, is called the *polynomial ring* over F. An element of $F[x]$ is called a *polynomial* over F. For a polynomial

$f(x) = \sum_{i=0}^{n} a_i x^i$, providing that $a_n \neq 0$, the integer n is called the *degree* of $f(x)$, denoted by $\deg(f(x))$. We define $\deg(0) = -\infty$. A nonzero polynomial $f(x)$ of degree n is said to be *monic* if $a_n = 1$. Furthermore, a polynomial $f(x)$ is said to be *reducible* over F if there exist two polynomials $g(x)$ and $h(x)$ over F such that $\deg(g(x)) < \deg(f(x))$ and $\deg(h(x)) < \deg(f(x))$, and $f(x) = g(x)h(x)$. A polynomial is said to be *irreducible* over F if it is not reducible..

§

Definition 34. Let $f(x)$ in $F[x]$ be a polynomial of degree $n \geq 1$. Then, for any polynomial $g(x)$ in $F[x]$ there exists a unique pair $(s(x), r(x))$ of polynomials, where $\deg(r(x)) < \deg(f(x))$ or $r(x) = 0$, such that $g(x) = s(x)f(x) + r(x)$. Here $r(x)$ is called the *principal remainder* of $g(x)$ divided by $f(x)$, or in our notation $(g(x) (\mathrm{mod}\, f(x)))$.

§

Definition 35. Let $f(x)$ and $g(x)$ in $F[x]$ be two nonzero polynomials. The *greatest common divisor* of $f(x)$ and $g(x)$, written $\gcd(f(x), g(x))$, is the monic polynomial of the highest degree which is a divisor of both $f(x)$ and $g(x)$. Two polynomials $f(x)$ and $g(x)$ are said to be *co-prime*, or *prime*, to each other if $\gcd(f(x), g(x)) = 1$. The *least common multiple* of $f(x)$ and $g(x)$, namely $\mathrm{lcm}(f(x), g(x))$, is the monic polynomial of the lowest degree which is a multiple of both $f(x)$ and $g(x)$.

§

Example 19. Let the factorisations of two polynomials $f(x)$ and $g(x)$ are,

$$f(x) = a \cdot (p_1(x))^{e_1} \cdots (p_n(x))^{e_n}$$

and

$$g(x) = b \cdot (p_1(x))^{d_1} \cdots (p_n(x))^{d_n}$$

where a and b are in F^*, and $e_i, d_i \geq 0$, and $p_i(x)$ are distinct monic irreducible polynomials, then,

$$\gcd(f(x), g(x)) = (p_1(x))^{\min(e_1, d_1)} \cdots (p_n(x))^{\min(e_n, d_n)}$$

and

$$\mathrm{lcm}(f(x), g(x)) = (p_1(x))^{\max(e_1, d_1)} \cdots (p_n(x))^{\max(e_n, d_n)}$$

Example 20. Let $f(x)$ and $g(x)$ in $F[x]$ be two nonzero polynomials. Then, there exist two polynomials $u(x)$ and $v(x)$ having $\deg(u(x)) < \deg(g(x))$ and $\deg(v(x)) < \deg(f(x))$ such that,

$$\gcd(f(x), g(x)) = u(x)f(x) + v(x)g(x)$$

Then,

$$\gcd(f(x)h(x), g(x)) = \gcd(f(x), g(x))$$

if $\gcd(h(x), g(x)) = 1$.

Theorem 22. Let $f(x)$ be a polynomial of degree n over a field F, where $n \geq 1$. Then $F[x]/(f(x))$, together with the addition,

$$g(x) \oplus h(x) = (g(x) + h(x) (\mathrm{mod}\, f(x)))$$

also written $g(x) + h(x)$, and multiplication,

$$g(x) \odot h(x) = (g(x)h(x) (\mathrm{mod}\, f(x)))$$

also written $g(x) \cdot h(x)$, form a ring. Furthermore, $F[x]/(f(x))$ is a field if and only if $f(x)$ is irreducible.

§

Exercise 9. Prove Theorem 22.

§

Example 21. Consider the ring $\mathbf{Z}_2[x]/(1+x^2) = \{0, 1, x, 1+x\}$ Its addition and multiplication tables are shown in Table 2.

$+$	0	1	x	$(1+x)$
0	0	1	x	$1+x$
1	1	0	$1+x$	x
x	x	$1+x$	0	1
$1+x$	$1+x$	x	1	0

$\times$	0	1	x	$1+x$
0	0	0	0	0
1	0	1	x	$1+x$
x	0	x	1	$1+x$
$1+x$	0	$1+x$	$1+x$	0

Table 2 *Addition and multiplication tables for* $\mathbf{Z}_2[x]/(1+x^2)$.

Example 22. Consider the ring $\mathbf{Z}_2[x]/(1+x+x^2)$. Its addition and multiplication tables are given in Table 3.

$+$	0	1	x	$1+x$
0	0	1	x	$1+x$
1	1	0	$1+x$	x
x	x	$1+x$	0	1
$1+x$	$1+x$	x	1	0

$\times$	0	1	x	$1+x$
0	0	0	0	0
1	0	1	x	$1+x$
x	0	x	$1+x$	1
$1+x$	0	$1+x$	1	x

Table 3 *Addition and multiplication tables for* $\mathbf{Z}_2[x]/(1+x+x^2)$.

Example 23. Table 4 shows the analogies between $\mathbf{Z}$ and $F[x]$.

the integral ring $\mathbf{Z}$	the polynomial ring $F[x]$
an integer m	a polynomial $f(x)$
a prime number p	an irreducible polynomial $p(x)$
$\mathbf{Z}_m = \{0, \ldots, m-1\}$	$F[x]/(f(x)) = \{\sum_{i=0}^{n-1} a_i x^i; a_i \in F, n \geq 1\}$
$a \oplus b = (a + b (\mathrm{mod}\, m))$	$g(x) \oplus h(x) = (g(x) \oplus h(x) (\mathrm{mod}\, f(x)))$
$a \odot b = (ab (\mathrm{mod}\, m))$	$g(x) \odot h(x) = (g(x)h(x) (\mathrm{mod}\, f(x)))$
$\mathbf{Z}_m$ is a ring	$F[x]/(f(x))$ is a ring
$\mathbf{Z}_m$ is a field if and only if m is a prime	$F[x]/(f(x))$ is a field if and only if $f(x)$ is irreducible

Table 4 *Analogies between* $\mathbf{Z}$ *and* $F[x]$.

Theorem 23. For every element ϕ of a finite field F with n elements, $\phi^n = \phi$.

Proof. The case when $\phi = 0$ is trivial. Next, if $\phi \neq 0$, then we could list all the nonzero elements of F as $F^* = \{\phi_1, \ldots, \phi_{n-1}\}$. And since F is closed, we could multiply each element in F^* to obtain $F^* = \{\phi\phi_1, \ldots, \phi\phi_{n-1}\}$. Therefore $\phi_1 \cdots \phi_{n-1} = (\phi\phi_1) \cdots (\phi\phi_{n-1})$, which leads to $\phi^{n-1} = 1$. ¶

Corollary 23[1]. Let F be a subfield of E, and let $|F| = n$. Then an element ϕ of E is also in F if and only if $\phi^n = \phi$.

Proof. The *if* part was already proved in Theorem 23. For the *only if* part, if ϕ satisfy $\phi^n = \phi$, then it is a root of $x^n - x$. And since $|F| = n$ means that all the elements of F are roots of $x^n - x$, it follows that ϕ lies in F. ¶

Definition 36. We denote a finite field with q elements by $\mathbf{F}_q$ or $GF(q)$. Let α be a root of an irreducible polynomial $f(x)$ of degree n over a field F. Then, if we replace x in $F[x]/(f(x))$ by α, the field $F[x]/(f(x))$ can be represented as,

$$F[\alpha] = \left\{ \sum_{i=0}^{n-1} a_i \alpha^i \right\}$$

for a_i in F.

§

Definition 37. An element α in a finite field $\mathbf{F}_q$ is called a *primitive element*, or *generator*, of $\mathbf{F}_q$ if $\mathbf{F}_q = \{0, \alpha, \alpha^2, \ldots, \alpha^{q-1}\}$.

§

Definition 38. The *order*, $\mathrm{ord}(\alpha)$, of a nonzero element α in $\mathbf{F}_q$ is the smallest positive integer k such that $\alpha^k = 1$.

§

Bibliography

George F Simmons. *Topology and modern analysis.* McGraw-Hill, 1963

L R Vermani. *Elements of algebraic coding theory.* Chapman & Hall, 1996

Exercises for Group, field and finite field

14^{th} January, 2007

15. Write the addition and multiplication tables for $GF(5) = \{0, 1, 2, 3, 4\}$
16. Find the principal remainder when $83 \cdot 54$ is divided by 7.
17. Determine whether $\left(3^{18}\right)\left(13^{35}\right) + 1$ is divisible by 17
18. Prove that $1+x+x^3$ and $1+x^2+x^3$ are the only irreducible polynomials of degree 3 over $\mathbf{F}_2$.
19. Is $GF(4)$ a subfield of $GF(8)$? Explain.
20. Construct the addition and multiplication tables for the rings $\mathbf{Z}_8$.
21. Find the multiplicative inverse of 3, 6, 10 in $\mathbf{Z}_{11}$.
22. Show that the polynomials $1+x^2$ and $2+2x+x^2$ over $\mathbf{F}_3$ are irreducible.
23. Factorise the polynomials $x^7 - 1$ over $\mathbf{F}_3$, $x^{20} - 1$ over $\mathbf{F}_7$ and $x^{11} - 1$ over $\mathbf{F}_5$.
24. Determine the number of primitive elements in the fields $\mathbf{F}_{10}$, $\mathbf{F}_{11}$ and $\mathbf{F}_{30}$.
25. Find the number of monic irreducible cubic polynomials over $\mathbf{F}_q$.
26. Find all the cyclotomic cosets of 2 modulo 33.
27. Let

$$f(x) = \left(2 + 2x^2\right)\left(1 + x^2 + x^3\right)^2\left(-1 + x^5\right)$$

in $\mathbf{F}_3[x]$ and

$$g(x) = \left(1 + x^2\right)\left(-2 + 2x^2\right)\left(1 + x^2 + x^3\right)$$

in $\mathbf{F}_3[x]$. Find $\gcd\left(f(x), g(x)\right)$ and $\operatorname{lcm}\left(f(x), g(x)\right)$.
28. Find two polynomials $u(x)$ and $v(x)$ in $\mathbf{F}_2[x]$ such that $\deg\left(u(x)\right) < 5$, $\deg\left(v(x)\right) < 4$ and

$$u(x)\left(1 + x + x^3\right) + v(x)\left(1 + x + x^2 + x^3 + x^4\right) = 1$$

29. Construct the addition and multiplication tables for the ring

$$\mathbf{F}_3[x]/\left(x^2 + 1\right)$$

30. Determine all the subfields in $\mathbf{F}_{2^{13}}$.

Reference

Raymond Hill. *A first course in coding theory.* Clarendon, 1986

San Ling and Chaoping Xing. *Coding theory, a first course.* Cambridge University Press, 2004

L R Vermani. *Elements of algebraic codng theory.* Chapman & Hall, 1996

Bounds in coding

18^{th} November 2005

Definition 39. A *prime power* is a prime or an integer power of a prime.

§

Example 24. Examples of prime powers are,

$$2, 3, 4, 5, 7, 8, 9, 11, 13, 16, 17, \ldots$$

Definition 40. Let the alphabet be $\mathbf{F}_q$, in other words a Galois field $GF(q)$, where q is a prime power, and let the vector space $V(n, q)$ be $(\mathbf{F}_q)^n$. Then a *linear code* over $GF(q)$, for some positive integer n, is a subspace of $V(n, q)$.

§

Theorem 24. A subset C of $V(n, q)$ is a linear code if and only if,

a. $\mathbf{u} + \mathbf{v} \in C$ for all $\mathbf{u}$ and $\mathbf{v}$ in C
b. $a\mathbf{u} \in C$ for all $\mathbf{u} \in C$ and $a \in GF(q)$

Proof. The proof follows from Definition 40 since, if C is a field, it must be closed under addition and multiplication. ¶

Example 25. A binary code is linear if and only if the sum of any two code words is a code word.

Definition 41. A *vector space* V is a set which is closed under finite vector addition and scalar multiplication. If the scalars are members of a field F, then V is called a vector space under F. Furthermore, V is a vector space under F if and only if for all members of V and F the following properties hold under addition,

a. commutativity
b. associativity
c. existence of an identity
d. existence of an inverse

while under multiplication the following,

e. associativity under scalar multiplication
f. distributivity of scalar sum
g. distributivity of vector sum
h. existence of a scalar multiplication identity

In other words, for all $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ in V and all p and q in F,

a. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$
b. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$
c. $\mathbf{0} + \mathbf{x} = \mathbf{x} + \mathbf{0} = \mathbf{x}$
d. $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$
e. $r(s\mathbf{x}) = (rs)\mathbf{x}$
f. $(r + s)\mathbf{x} = r\mathbf{x} + s\mathbf{x}$
g. $r(\mathbf{x} + \mathbf{y}) = r\mathbf{x} + r\mathbf{y}$
h. $1\mathbf{x} = \mathbf{x}$

§

Example 26. Let q be a prime power, and let $GF(q)$ denote a finite field over q elements. Then, by *vector space over finite field* we mean a set $GF(q)^n$ of all ordered n-tuples over $GF(q)$, which is closed under finite vector addition and multiplication, that is to say, multiplication by some scalar a in $GF(q)$.

Theorem 25. A non-empty subset C of $V(n,q)$ is a subspace if and only if C is closed under addition and scalar multiplication. In other words

Proof. What Theorem 25 states amounts to saying that a non-empty C in $V(n,q)$ is a subspace if and only if,
a. $\mathbf{x}, \mathbf{y} \in C$ implies $\mathbf{x} + \mathbf{y} \in C$
b. if $a \in GF(q)$ and $\mathbf{x} \in C$, then $a\mathbf{x} \in C$
All properties to be met in Definition 41 are the same for C as for $V(n,q)$ itself, provided that C is closed under addition and scalar multiplication. Therefore statements (a) and (b) are necessary for C to be a subspace. They are also sufficient since C is already a subset of $V(n,q)$. ¶

Definition 42. A *linear combination* of r vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$ in $V(n,q)$ is any vector of the form $\sum_{i=1}^{r} a_i \mathbf{v}_i$, where a_i are scalars. Let A be a set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$. Then A is said to be *linearly dependent* if there exist scalars $a_1, \ldots, a_r$ not all of which are zero, such that $\sum_{i=1}^{r} a_i \mathbf{v}_i = \mathbf{0}$. And A is *linearly independent* if it is not linearly dependent, that is to say, if $\sum_{i=1}^{r} a_i \mathbf{v}_i = \mathbf{0}$ implies a_i are all zero for $i = 1, \ldots, r$.

§

Definition 43. Let C be a subspace of a vector space $V(n,q)$ over $GF(q)$. Then a subset $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ of C is called a *generating-* or *spanning set* of C if every vector in C can be expressed as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_r$. A *basis* of C is a generating set of the same which is also linearly independent.

§

Definition 44. For a q-ary (n,m,d)-code C, the *relative minimum distance* of C is defined to be

$$\delta(C) = \frac{d-1}{n}$$

§

Definition 45. Let a code alphabet A be of size $q > 1$, n the size of each word, d the minimum distance, and $A_q(n,d)$ the largest possible vocabulary size m such that there exists an (n,m,d)-code over A. Then any (n,m,d)-code C which has $m = A_q(n,d)$ is called an *optimal code*. The *main coding theory problem* is precisely to find the value of $A_q(n,d)$.

§

Definition 46. Consider each word as an n-tuple. Then all such tuples lying within Hamming distance r of an n-tuple x are said to be within a *Hamming sphere* of radius r around x.

§

Theorem 26. Let the size of the alphabet be $q = |A|$, the size of a word be n, and the Hamming- or minimum distance be d. Then the Hamming- or sphere-packing bound on the size m of a code dictionary C is given by,

$$m \leq \frac{q^n}{\sum_{i=0}^{r^90} (q-1)^i \binom{n}{i}}$$

where

$$r^90 = \left\lfloor \frac{d-1}{2} \right\rfloor$$

Proof. Let c be a code word. Let $e(x, y)$ be the number of places which are different between two words x and y. Since there are $q-1$ possibilities for each differing position between any two words, there are $(q-1)^i$ possible errors when i places are different. And to position these i places there are altogether $\binom{n}{i}$ ways. Therefore the number of all words w_i such that $e(w_i, c) \leq r$ is the number n_r of n-tuples in a Hamming sphere of radius r around c, and is,

$$n_r = \sum_{i=0}^{r} (q-1)^i \binom{n}{i}$$

Then the lower bound for our code is $d(C) > 2r$, that is to say, $d(C) \geq 2r+1$. In other words, Hamming spheres of radius r around the m code words of C are mutually nonintersecting. There are a total of q^n possible n-tuples, that is words of length n, not all of which are code words. In other words, $m < q^n$. And since there are n_r of these n-tuples within each sphere, the the number of the all the n-tuples contained within the space of all these n-tuples over the alphabet A is $n_r m$. Hence,

$$m \sum_{i=0}^{r} (q-1)^i \binom{n}{i} \leq q^n$$

and thus this theorem. ¶

Definition 47. Codes which satisfies the Hamming bound are called [perfect codes].

§

Problem 1. Let r and n be integers such that $0 < r < \frac{n}{2}$, then prove that,

$$\left[8n\left(\frac{r}{n}\right)\left(1-\frac{r}{n}\right)\right]^{-\frac{1}{2}} 2^{nH\left(\frac{r}{n}, 1-\frac{r}{n}\right)} \leq \sum_{i=0}^{r} \binom{n}{i} \leq 2^{nH\left(\frac{r}{n}, 1-\frac{r}{n}\right)}$$

where $H(x, y)$ is the entropy function the arguments x and y of which are probabilities and $H(\cdot, \cdot)$ has the unit of bits per symbol. (Hint: Stirling's approximation to $n!$, *cf* MacWilliams and Sloane, 1977)

§

Note 1. Let $C(n,d)$ be a code with words of length n and minimum distance between words d. Let $m_{n,d}$ be the number of code words in $C(n,d)$. Then the size of the largest dictionary of n-tuples with fractional minimum distance d_f is,

$$m_m(n,d_f) = \max_{\left\{C(n,d):\left(\frac{d}{n}\right)\geq d_f\right\}} |C(n,d)|$$

§

Problem 2. From Note 1, show that for n fixed, $m_m(n,d_f)$ is a monotonous nonincreasing function of d_f. Then show that with d_f fixed, $m_m(n,d_f)$ increases exponentially with n.

§

Definition 48. The *asymptotic transmission rate* is defined to be,

$$R(d_f) = \lim_{n\to\infty} \frac{1}{n} \log m_m(n,d_f)$$

Also defined are the upper- and the lower bounds on this rate,

$$\bar{R}(d_f) = \limsup_{n\to\infty} \frac{1}{n} \log m_m(n,d_f)$$

and

$$\underline{R}(d_f) = \liminf_{n\to\infty} \frac{1}{n} \log m_m(n,d_f)$$

§

Note 2. For large n, show that $\underline{R}\,(d_f) < R\,(d_f) < \bar{R}\,(d_f)$.

§

Example 27. Using the results from Problem 1 we obtain the Hamming bound for the binary code,

$$m \leq \left(8n\left(\frac{r}{n}\right)\left(1-\frac{r}{n}\right)\right)^{\frac{1}{2}} 2^{n\left(1-H\left(\frac{r}{n},1-\frac{r}{n}\right)\right)} \tag{16}$$

where

$$r = \left\lfloor \frac{d-1}{2} \right\rfloor$$

Equation 16 must hold for all binary dictionaries, therefore it gives an upper bound on the maximum dictionary size $m_m(n,d_f)$ over all dictionaries whose word length is n and fractional distance,

$$d_f = \frac{d}{n} = \frac{2r + \left\{\begin{matrix}1\\2\end{matrix}\right\}}{n}$$

where the choice of 1 or 2 depends on whether d is odd or respectively even. For large n,

$$m_m(n, d_f) \leq \left(9n\left(\frac{d_f}{2}\right)\left(1-\frac{d_f}{2}\right)\right)^{\frac{1}{2}} 2^{n\left(1-H\left(\frac{d_f}{2}, 1-\frac{d_f}{2}\right)\right)}$$

The upper bound for the attainable information rate is,

$$\begin{aligned}
&\bar{R}(d_f) \\
&\quad = \limsup_{n\to\infty} \frac{1}{n} \log_2 m_m(n, d_f) \\
&\quad \leq \lim_{n\to\infty} \left\{ \frac{1}{2} \frac{\log_2 n}{n} + \frac{1}{2n} \log_2 \left(\frac{9d_f}{2}\left(1-\frac{d_f}{2}\right)\right)\right\} + 1 - H\left(\frac{d_f}{2}, 1-\frac{d_f}{2}\right)
\end{aligned}$$

As n approaches infinity,

$$\bar{R}(d_f) \leq 1 - H\left(\frac{d_f}{2}, 1-\frac{d_f}{2}\right)$$

Problem 3. Work out the details of derivation of Example 27.

§

Theorem 27. Let $d(c_i, c_j)$ be the Hamming distance between the code words c_i and c_j. Let $d(C)$ be the minimum distance between code words, and $\bar{d}$ the average distance between words. If,

$$\frac{d}{n} > \frac{q-1}{q}$$

then the *Plotkin's bound*,

$$m_{n,d} \leq \frac{\frac{d}{n}}{\frac{d}{n} - \frac{q-1}{q}}$$

Proof. The average distance gives an upper bound for the minimum distance, that is $d \leq \bar{d}$, where

$$\begin{aligned}
\bar{d} &= \frac{\sum_{i=2}^{m} \sum_{j=1}^{i-1} d\left(c_i, c_j\right)}{\sum_{i=2}^{m} \sum_{j=1}^{i-1} 1} \\
&= \left(\frac{m(m-1)}{2}\right)^{-1} \sum_{i=2}^{m} \sum_{j=1}^{i-1} d\left(c_i, c_j\right)
\end{aligned}$$

Since the Plotkin's bound is an upper bound on d, we need to maximise,

$$\begin{aligned}
\sum_{i>j} \sum d\left(c_i, c_j\right) &= \sum_{i>j} \sum \sum_{k=1}^{n} d\left(c_{ik}, c_{jk}\right) \\
&= \sum_{k=1}^{n} \sum_{i>j} \sum d\left(c_{ik}, c_{jk}\right)
\end{aligned}$$

This implies (*cf* Plotkin, 1960),

$$\sum_{i>j}\sum d\left(c_i, c_j\right) \leq \sum_{k=1}^{n} \max_{\{c_{ik}, i=1,\ldots,m\}} \left\{ \sum_{i>j}\sum d\left(c_{ik}, c_{jk}\right) \right\}$$

which says that the upper bound is maximised by choosing a maximising c_{ik} from the alphabet A. However this is,

$$\max_{c_{ik}, i=1,\ldots,m} \sum_{i>j}\sum d\left(c_{ik}, c_{jk}\right) \leq \left(\frac{m}{q}\right)^2 \frac{q(q-1)}{2}$$

Providing that,

$$\frac{d}{n} > \frac{q-1}{q}$$

then

$$d \leq n\left(\frac{m}{m-1}\right)\left(\frac{q-1}{q}\right)$$

¶

Note 3. Notice how,

$$\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}(\cdot) = \sum_{i=2}^{m}\sum_{j=1}^{i-1}(\cdot)$$

Equivalently to this are,

$$\sum_{i<j}\sum(\cdot) \quad \text{and} \quad \sum_{i>j}\sum(\cdot)$$

§

Problem 4. Prove Note 3 on double summations.

§

Note 4. If,

$$d_f > \frac{q-1}{q}$$

then,

$$m_m(n, d_f) \leq \frac{d_f}{d_f - \left(\frac{q-1}{q}\right)}$$

and then,

$$\bar{R}(d_f) = \limsup_{n\to\infty} \frac{1}{n} \log m_m(n, d_f) = 0$$

On the other hand if,

$$d_f \leq \frac{q-1}{q}$$

then from,

$$m(n,d) = \sum_{a \in A} m_a(n,d)$$

where $m(n,d) = |C(n,d)|$, $C(n,d)$ being any code consisting of n-tuples whose minimum distance is at least d, and $m_x(n,d) = |C_x(n,d)|$, $C_x(n,d)$ comprising all n-tuples in $C(n,d)$ which begin with the symbol x. Hence,

$$\begin{aligned} m(n,d) &\leq q m_x(n,d) \\ &= q m(n-1,d) \\ &\vdots \\ &= q^{n-k} m(k,d) \end{aligned}$$

Provided k is small enough, we may yet use the Plotkin's bound, hence

$$m(n,d) \leq \frac{q^{n-k}\left(\frac{d}{k}\right)}{\left(\frac{d}{k}\right) - \left(\frac{q-1}{q}\right)}$$

when

$$\frac{d}{k} > \frac{q-1}{q}$$

Choose k the largest integer satisfying

$$\frac{d}{k} - \frac{1}{qk} \geq \frac{q-1}{q}$$

Then,

$$k + r = \frac{qd-1}{q-1}$$

where $0 \leq r < 1$. And then,

$$m(n,d) \leq \frac{q^{n-\left(\frac{qd-1}{q-1}\right)r+1} d}{(q-1)r+1}$$

Finally,

$$m(n,d) \leq q^{n-\left(\frac{qd-1}{q-1}\right)} d$$

and, if d_f is fixed and n become large,

$$\bar{R}(d_f) \leq \log q \left(1 - \frac{q}{q-1} d_f\right)$$

§

Problem 5. Prove that,

$$q^r \left((q-1)\, r + 1\right)^{-1} \leq 1$$

for $0 \leq r \leq 1$

§

Proposition 1. Let C be a code containing binary n-tuples, $m_d(x)$ the number of code words within distance d of an n-tuple x. Further, let A be a new code whose code words are the difference vectors $a_1, \ldots, a_{m_d}$ such that $a_i = c_i \ominus x$, $i = 1, \ldots, m_d$, where $\ominus$ denotes modulo subtraction of the vectors, element by element. Assume that $d < \frac{n}{2}$ and both d and m are large enough such that $m_d(x) \geq 2$. Then,

$$\frac{d_c}{n} \leq \frac{2d}{n}\left(1 - \frac{d}{n}\right)\frac{m_a}{m_a - 1} \tag{17}$$

where

$$m_a \geq \left\lceil 2^{-n} m \sum_{i=0}^{d} \binom{n}{i} \right\rceil$$

Proof. Since C is a code of binary n-tuples, there are

$$\sum_{i=0}^{d} \binom{n}{i}$$

n-tuples within distance d of each code word. This gives the total of

$$m \sum_{i=0}^{d} \binom{n}{i}$$

n-tuples in the Hamming sphere around the m code words.
There are $m_d(x)$ code words within the distance d of any n-tuple x. For x in X^n, c in C and $d(x, c) \leq d$, the number of pairs (x, c) can be counted by picking up first x and then c, hence

$$\sum_{x \in X^n} m_d(x) = m \sum_{i=0}^{d} \binom{n}{i}$$

Since X^n contains 2^n of n-tuples, consequently there exists some value of x such that,

$$m_d(x) \geq \left\lceil 2^{-n} m \sum_{i=0}^{d} \binom{n}{i} \right\rceil$$

Let $c_1, \ldots, c_{m_d}$ be code words in C that lie within Hamming distance d of the n-tuple x. Consider the difference vector $a_1, \ldots, a_{m_d}$ such that $a_i = c_i \ominus x$. Then A is a set of localised code words of C. Then,

$$a_i \ominus a_j = (c_i \ominus x) \ominus (c_j \ominus x) = c_i \ominus c_j$$

and we have,

$$d(c_i, c_j) = d(a_i, a_j)$$

Thus,

$$m_a \geq m_d(x) \geq \left\lceil 2^{-n} m \sum_{i=0}^{d} \binom{n}{i} \right\rceil$$

Also, $d_a \geq d_c$ and $w(a_i) \leq d$ for all n-tuple a_i in A, where the Hamming weight $w(a_i)$ is the number of nonzero elements in a_i.
Next, applying the average-distance Plotkin bound to the localised code A one obtains,

$$d_c \leq d_a \leq \bar{d}_a = \left(\frac{m_a(m_a - 1)}{2} \right)^{-1} \sum_{i>j} \sum d(a_i, a_j) \tag{18}$$

We maximise RHS of Equation 18 to get rid of the dependence on A. We enlarge our restriction on $w(a_i)$ above to the set of all possible a_i in A, thus,

$$\sum_{a_i \in A} w(a_i) \leq m_a d \tag{19}$$

Then, let z_k be the number of code words in A having a 0 in the k^{th} position. We maximise,

$$\sum_{i>j} \sum d(a_i, a_j) = \sum_{k=1}^{n} (m_a - z^k) \tag{20}$$

subject to the constraint of Equation 19 that,

$$\sum_{k=1}^{n} (m_a - z_k) \leq m_a d \tag{21}$$

By setting,

$$z_k = \frac{m_a d}{n} \tag{22}$$

we maximise RHS of Equation 20 under the constraint in Equation 21. From Equation's 18, 20 and 22 we obtain Equation 17. ¶

Algorithm 2 *Gilbert bound, a lower bound to m for n, d and q.*

```
S^n ← X^n
for all c_i in S^n do
   for all n-tuples c_j within d − 1 distance of C do
      remove c_j
   endfor
endfor
```

Note 5. For the Gilbert bound algorithm, Algorithm 2, initially $|S|^n = |X|^n$. For each c_i chosen, at most

$$\sum_{i=0}^{d-1}(q-1)^i \binom{n}{i}$$

n-tuples are removed. If

$$(m-1)\sum_{i=0}^{d-1}(q-1)^i \binom{n}{i} < q^n$$

then the algorithm will not stop after $m-1$ code-word selections.

§

Bibliography

Solomon W Golomb, Robert E Peile and Robert A Scholtz. *Basic concepts in information theory and coding, the adventures of secret agent 00111.* Plenum Press, 1994

Raymond Hill. *A first course in coding theory.* Clarendon, 1986

San Ling and Chaoping Xing. *Coding theory, a first course.* Cambridge University Press, 2004

M Plotkin. Binary codes with specified minimum distance. *IRE Transaction on Information Theory.* **6**, 445–50, 1960

F J MacWilliams and N J A Sloane. *The theory of error-correcting codes.* North-Holland, Amsterdam, 1977

Group-, polynomial-, and Hamming codes

25^{th} November 2005

A recapitulation of group, ring, field and finite field is given in Definition 49 and Example 28.

Definition 49. A non-empty set G with a binary composition is called a *group* if the composition is associative, if a unique *identity* exists for all elements in G, and if a unique *inverse* exists for each of the elements in G. The group G is called *Abelian* if the composition in it is commutative for any two elements in G. A non-empty set R with two binary compositions, call these addition and multiplication, defined on it is called a *ring* if R is an Abelian group with respect to the composition addition, if multiplication in R is associative, and if distributive laws hold for all elements in R. A set F having at least two elements with two compositions, be them called addition and multiplication, defined on it is called a *field* if it is a commutative ring with identity every non-zero element of which has an inverse with respect to multiplication. A field having only a finite number of elements is called a *finite* or *Galois field.*

§

Example 28. The set $F_p = \{0, \ldots, p-1\}$ in which addition and multiplication are defined modulo p, where p is a prime integer, is a finite field. For $p = 2$ we have $F_2 = \{0, 1\}$, which is denoted by $\mathbf{B}$. The set $\mathbf{B}^n$ of all ordered n-tuples or sequences of length n, a positive integer, with each tuple or entry of the sequence being in the field $\mathbf{B}$ and a composition defined as a componentwise summation of any two sequences in $\mathbf{B}^n$, is an Abelian group. The zero sequence of length n is the identity of $\mathbf{B}^n$ and each element in $\mathbf{B}^n$ is its own inverse.

Definition 50. A *binary block* (b, n)*-code* comprises an *encoding function* $E : \mathbf{B}^b \to \mathbf{B}^n$ and a *decoding function* $D : \mathbf{B}^n \to \mathbf{B}^b$. The images of E are called *code words.*

§

Definition 51. Let two binary sequences be a and b in $\mathbf{B}^n$. The *distance* $d(a, b)$ between a and b is defined as

$$d(a, b) = \sum_{i=1}^{n} x_i$$

where

$$x_i = \begin{cases} 0 & \text{if } a_i = b_i \\ 1 & \text{if } a_i \neq b_i \end{cases}$$

§

Definition 52. The *weight* $w(a)$ of a in $\mathbf{B}^n$ is the number of non-zero components of the sequence a.

§

Theorem 28. Let a and b be any two sequences in $\mathbf{B}^n$. Then $d(a,b) = w(a+b)$.

Proof. The only contribution of 1 to $d(a,b)$ is $a_i \neq b_i$ for all $1 \leq i \leq n$. But this latter is the case if and only if $a_i + b_i = 1$, and this contributes 1 to $w(a+b)$. ¶

Definition 53 recapitulates the concept of homomorphism, whereas Definition 54 defines what a group code is.

Definition 53. Let X and Y be two groups. Then a map $f : X \to Y$ which satisfies the property $f(x_1x_2) = f(x_1)f(x_2)$ for all x_1 and x_2 in X is called a *homomorphism*. Further, the homomorphism f is called a *monomorphism* if it is one to one, and it is called an *isomorphism* if it is both one to one and onto.

§

Definition 54. A block code is called a *group code* if all its code words form an additive group.

§

Definition 55. A $b \times n$ matrix G over $\mathbf{B}$, where $b < n$, is called an *encoding-* or *generator matrix* if G is of the form

$$G = [I_b \, G_n]$$

where I_b is an identity matrix of dimension b and G_n a $b \times (n-b)$ matrix. An *encoding function* $E : \mathbf{B}^b \to \mathbf{B}^n$ is defined by

$$E(x) = xG$$

for all x in $\mathbf{B}^b$

§

Theorem 29. The encoding function $E : \mathbf{B}^b \to \mathbf{B}^n$ given by $E(x) = xG$ for all x in $\mathbf{B}^b$, where G is a $b \times n$ generator matrix, is a monomorphism.

Proof. Both $\mathbf{B}^b$ and $\mathbf{B}^n$ are additive Abelian groups. Then for all x and y in $\mathbf{B}^b$ we know that $x+y$ is also in $\mathbf{B}^b$ and $E(x+y) = (x+y)G = xG + yG = E(x) + E(y)$. Thus E is a homomorphism. Further, as the first part of G is I_b, it follows that a part of $E(x)$ is x itself. Therefore the matrix encoding method gives for each binary message word a distinct code word. In other words, the mapping E is one to one, which means that it is a monomorphism. ¶

Definition 56. A code generated by a generating matrix is called a *matrix code.*

§

Theorem 30. A matrix code is a group code.

Proof. The code words generated by E are associative, since

$$x_1G + (x_2G + x_3G) = (x_1G + x_2G) + x_3G$$

They have a unique identity, that is the zero $b \times n$ matrix, and each of them is its own inverse. ¶

Definition 57. An $(b, b+1)$ parity check code is the code generated by an encoding function $E : \mathbf{B}^b \to \mathbf{B}^{b+1}$ defined by

$$E(a_1 \cdots a_b) = a_1 \cdots a_b a_{b+1}$$

where

$$a_{b+1} = \begin{cases} 1 & \text{if } w(a) \text{ is odd} \\ 0 & \text{if } w(a) \text{ is even} \end{cases}$$

$w(a)$ being $w(a_1 \cdots a_b)$.

§

Theorem 31. The $(b, b+1)$ parity check code is a group code.

Proof. Let our unencoded binary words be $a = a_1 \cdots a_b$, $b = b_1 \cdots b_b$, and $c = c_1 \cdots c_b$ such that $c_i = a_i + b_i$ for $i = 1, \ldots, b$, and let the coded words of a and b be respectively $\bar{a} = aa_{b+1}$ and $\bar{b} = bb_{b+1}$. Since c is odd if and only if either a is odd while b is even or vice versa, but when this is the case we have either $a_{b+1} = 1$ and $b_{b+1} = 0$, or $a_{b+1} = 0$ and $b_{b+1} = 1$. Either way we have $c_{b+1} = 1 = a_{b+1} + b_{b+1}$. Next, c is even if and only if a and b are either both odd or both even. But when either of these is the case, then $a_{b+1} + b_{b+1} = 0 = c_{b+1}$. Hence $\bar{c}$ is a parity-check code word. The zero word is the identity and the inverse of each word is that word itself. Therefore the set of all code words forms a group. ¶

Theorem 32. The minimum distance of a group code equals the minimum of the weights of its non-zero code words.

Proof. Let d_m be the minimum distance of the group code, and w_m the minimum of the weights of the non-zero code words of the same. Then there exist code words a and b such that $d_m = d(a, b) = w(a + b) \geq w_m$. Now, w_m implies that there exists a non-zero code word c such that $w_m = w(c) = d(c, 0) \geq d_m$. Hence $d_m = w_m$. ¶

Example 29. Let the generator matrix be

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

The dimension of G is $b \times n$, which in this case is 3×6. Let $a_1a_2a_3a_4a_5a_6$ be the code word and $a_1a_2a_3$ the original word, then

$$(a_1\, a_2\, a_3\, a_4\, a_5\, a_6) = (a_1\, a_2\, a_3)\, G$$

and then,

$$\begin{aligned} a_4 &= a_1 + a_2 \\ a_5 &= a_1 + a_3 \\ a_6 &= a_1 + a_2 + a_3 \end{aligned}$$

In other words,

$$\left.\begin{aligned} a_1 + a_2 + a_4 &= 0 \\ a_1 + a_3 + a_5 &= 0 \\ a_1 + a_2 + a_3 + a_6 &= 0 \end{aligned}\right\} \text{parity check equations}$$

These parity check equations are then, in matrix form,

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{pmatrix} = 0$$

The matrix

$$H = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

is called the *parity check matrix* of the code. Then $G = (\, I_3 \quad A\,)$ and $H = (\, A^90 \quad I_3\,)$, where

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

and

$$A^90 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

Note 6. Let $\mathbf{a}$ be the vector formed from elements of the code word a. Then we have that, for any code word a the parity check matrix H has the property that $H\mathbf{a} = 0$.

§

Example 30. The parity check code in Definition 57 is in fact a matrix code given by the generator matrix

$$G = \begin{pmatrix} 1 & 0 & \cdots & 0 & 1 \\ 0 & 1 & & 0 & 1 \\ \vdots & & \ddots & & \vdots \\ 0 & & \cdots & 1 & 1 \end{pmatrix}$$

whose parity check matrix is the $1 \times (b+1)$ matrix $H = (\,1 \quad \cdots \quad 1\,)$.

Definition 58. The *syndrome* of a word $r \in \mathbf{B}^n$ is $\mathbf{s} = H\mathbf{r}^9 0$.

§

In syndrome decoding algorithm of Algorithm 3, r is a received word, s its syndrome, b_r the decoded original word, while c_r the decoded code word.

Algorithm 3 *the syndrome decoding algorithm.*

$r \leftarrow r_1 \cdots r_b r_{b+1} \cdots r_n$
$\mathbf{s} \leftarrow H\mathbf{r}^9 0$
if $\mathbf{s} = 0$ **then**
  $b_r \leftarrow (r_1 \cdots r_b)$
elseif $\mathbf{s}$ matches the i^{th} column of H **then**
  $c_r \leftarrow (r_1 \cdots r_{i-1}(r_i + 1)r_{i+1} \cdots r_n)$
  $b_r \leftarrow (c_{r1} \cdots c_{rb})$
else
  at least two errors have occurred in the transmission
endif

Theorem 33. An $(n-b) \times b$ parity check matrix H will decode all single errors correctly if and only if the columns of H are distinct and non-zero.

Proof. Suppose the i^{th} column of H is zero, and let e be a word whose weight is 1 having 1 in the i^{th} position and 0 elsewhere. Then for any code word b, we have $H(\mathbf{b}+\mathbf{e})^9 0 = H\mathbf{b}^9 0 + H\mathbf{e}^9 0 = 0$. So our decoding procedure becomes $D(b+e) = b+e$ and the error vector $\mathbf{e}$ goes undetected.
Next, suppose that the i^{th} and the j^{th} columns of H are identical. Let e^i and e^j be words of length n with 1 in the i^{th} and respectively j^{th} position and 0 elsewhere. Then for any code word b, we have $H(\mathbf{b}+\mathbf{e}^i)^9 0 = H\mathbf{b}^9 0 + H\left(\mathbf{e}^i\right)^9 0 = H\left(\mathbf{e}^i\right)^9 0 = H\mathbf{b}^9 0 + H\left(\mathbf{e}^j\right)^9 0 = H\left(\mathbf{b}+\mathbf{e}^j\right)^9 0$. We are unable to decide whether the error occurred in the i^{th} or the j^{th} position.
Conversely, suppose all the columns of H are distinct and non-zero. Then for any code word b and any error vector $\mathbf{e}$ of weight 1 having 1 in the i^{th} position, $H\left(\mathbf{b}+\mathbf{e}\right)^9 0 = H\left(\mathbf{b}^9 0 + \mathbf{e}^9 0\right) = H\mathbf{b}^9 0 + H\mathbf{e}^9 0 = 0 + H\mathbf{e}^9 0$. Our decoding procedure gives $D(b+e) = b$, therefore every single error is corrected. ¶

Theorem 34. If $G = (\,I_b \quad A\,)$ is a $b \times n$ generator matrix of a code, then $H = (\,A^9 0 \quad I_{n-b}\,)$ is the unique parity check matrix for the same code. If

$H = (\, B \quad I_{n-b} \,)$ is an $(n-b) \times n$ parity check matrix, then $G = (\, I_m \quad B^9 0 \,)$ is the unique generator matrix for the same code.

Proof. Let the original word be $a \in \mathbf{B}^b$ and c be the code word corresponding to a with respect to the code given by the generator matrix G. Then $\mathbf{c} = \mathbf{a}G$ Let a be $a_1 \cdots a_b$. Since the first b columns of G is an identity matrix, it follows from $\mathbf{c} = \mathbf{a}G$ that $a_i = b_i$ for all $1 \leq i \leq b$. Let $\bar{c} = c_{b+1} \cdots c_n$, then $c = c_1 \cdots c_b c_{b+1} \cdots c_n$ and $\mathbf{c} = (\, \mathbf{a} \quad \bar{\mathbf{c}} \,)$. Then,

$$\begin{aligned}
H\mathbf{c}^9 0 &= (\, A^9 0 \quad I_{n-b} \,)\, (\mathbf{a}G)^9 0 \\
&= (\, A^9 0 \quad I_{n-b} \,)\, G^9 0 \mathbf{a}^9 0 \\
&= (\, A^9 0 \quad I_{n-b} \,)\, (I_m A)^9 0 \mathbf{a}^9 0 \\
&= (\, A^9 0 \quad I_{n-b} \,) \begin{pmatrix} I_m \\ A^9 0 \end{pmatrix} \mathbf{a}^9 0 \\
&= (\, A^9 0 I_m + I_{n-b} A^9 0 \,)\, \mathbf{a}^9 0 \\
&= (\, A^9 0 + A^9 0 \,)\, \mathbf{a}^9 0 \\
&= 0 \times \mathbf{a}^9 0 \\
&= 0
\end{aligned}$$

Therefore c is the code word corresponding to the original word a in the code given by the parity check matrix H.
Now, suppose first that c is the code word corresponding to the original word a as above in the code obtained from the parity check matrix $H = (\, A^9 0 \quad I_{n-b} \,)$. Then $c_i = a_i$ for all $1 \leq i \leq b$ and $H\mathbf{c}^9 0 = 0$. Let $\bar{c} = c_{b+1} \cdots c_n$. Then,

$$\begin{aligned}
H \begin{pmatrix} \mathbf{a} \\ \bar{\mathbf{c}}^9 0 \end{pmatrix} &= 0 \\
(\, A^9 0 \quad I_{n-b} \,) \begin{pmatrix} \mathbf{a} \\ \bar{\mathbf{c}}^9 0 \end{pmatrix} &= 0 \\
A^9 0 \mathbf{a}^9 0 + I_{n-b} \bar{\mathbf{c}}^9 0 &= 0
\end{aligned}$$

Therefore $\bar{c} = \mathbf{a}A$, and

$$\mathbf{c} = (\, \mathbf{a} \quad \bar{c} \,) = (\, \mathbf{a}I_m \quad \mathbf{a}A \,) = \mathbf{a}\,(\, I_m \quad A \,) = \mathbf{a}G$$

Hence c is the code word corresponding to the original word a in the code defined by the generator matrix G. So far we have proved that codes determined by G and H are identical.
Suppose that to $G = (\, I_m \quad A \,)$ corresponds another parity check matrix $H_1 = (\, B \quad I_{n-b} \,)$. Let e^i be the original word with 1 in the i^{th} position and 0 elsewhere. The corresponding code word is $\mathbf{e}^i G$, that is the i^{th} row of G, or we may write $\mathbf{e}^i G = (\, \mathbf{e}^i \quad \tilde{\mathbf{e}}^i \,)$, where $\tilde{\mathbf{e}}^i$ is the i^{th} row of A. Since H_1 is a

parity check matrix of the code defined by G, it follows that,

$$\begin{aligned} H_1 \left(\mathbf{e}^i \quad \tilde{\mathbf{e}}^i \right)^9 0 &= 0 \\ \left(B \quad I_{n-b} \right) \begin{pmatrix} \left(\mathbf{e}^i\right)^9 0 \\ \left(\tilde{\mathbf{e}}^i\right)^9 0 \end{pmatrix} &= 0 \\ B \left(\mathbf{e}^i\right)^9 0 + \left(\tilde{\mathbf{e}}^i\right)^9 0 &= 0 \end{aligned}$$

Therefore $\left(\tilde{\mathbf{e}}^i\right)^9 0$ matches the i^{th} column of B, or equivalently $\tilde{\mathbf{e}}^i$ matches the i^{th} row of $B^9 0$. Then the i^{th} row of A is identical to the i^{th} column of B. And this is true for all $1 \leq i \leq b$, so we have $B = A^9 0$ and therefore $H_1 = H$. Hence, to a given G there corresponds a unique $H = \left(A^9 0 \quad I_{n-b} \right)$. Similar argument also holds if we start with a parity check matrix H given. ¶

Definition 59. Let C be a (b, n) code obtained from the generator matrix $G = [I_b \, A]$. Then an $(n-b, n)$ matrix code defined by the parity check matrix $H = [A \, I_b]$ is called the *dual code* $C^{\perp}$ of C.

§

Definition 60. Two words x and y are said to be in the same coset if and only if $y = x + c$ for some code word c in C.

§

Theorem 35. Two words x and y in $\mathbf{B}^n$ are in the same coset of C if and only if they have the same syndrome.

Proof. By Definition 60 x and y are in the same coset if and ony if $y = x + c$ for some c in C, which in turn is true if and only if $x + y = c$ in C. Then it follows from this that,

$$\begin{aligned} H(\mathbf{x} + \mathbf{y})^9 0 &= 0 \\ H\left(\mathbf{x}^9 0 + \mathbf{y}^9 0\right) &= 0 \\ H\mathbf{x}^9 0 + H\mathbf{y}^9 0 &= 0 \\ H\mathbf{x}^9 0 &= H\mathbf{y}^9 0 \end{aligned}$$

¶

Next, we look at polynomial codes. Before we do so we look briefly at vector space again. Definition's 61 and 62 are respectively about vector space and linear-dependence, and then Theorem 36 is on isomorphic vector spaces.

Definition 61. Let F be a field. Then a non-empty set V is called a *vector space* over F if V and an addition form an Abelian group; for every a in F and v in V there is a uniquely defined element av in V such that for any v, v_1 and v_2 in V and any a and b in F, $a(v_1 + v_2) = av_1 + av_2$; $(a+b)v = av + bv$; $(ab)v = a(bv)$; and $1v = v$, 1 being the identity of F.

§

Definition 62. Let V be a vector space over a field F. Then a set $\{v_1, \ldots, v_n\}$ of elements v_i in V is said to be *linearly independent* if $a_1v_1 + \cdots + a_nv_n = 0$, for $a_1, \ldots, a_n$ in F, implies $a_1 = \cdots = a_n = 0$. A set $\{v_1, \ldots, v_n\}$ is called a *basis* of V if all its elements $v_1, \ldots, v_n$ in V are linearly independent over F and all elements in V may be expressed in the form $a_1v_1 + \cdots + a_nv_n$ where all a_i, $i = 1, \ldots, n$, are in F. Also V is said to be of *dimension* n over F, $\dim V = n$. A map $f : V \to W$ from one vector space to another, where V and W are vector spaces over the same field F, is called an *isomorphism* if the map f one to one and onto and, for all v, v_1 and v_2 in V and a in F, $f(v_1 + v_2) = f(v_1) + f(v_2)$ and $f(av) = af(v)$.

§

Theorem 36. Let two vector spaces V and W over the same field F have the same finite dimension. Then V and W are isomorphic.

Proof. Let $\dim V = \dim W = n$. Let $\{x_1, \ldots, x_n\}$ be a basis of V over F, and $\{y_1, \ldots, y_n\}$ a basis of W over F. Since all the elements of V can be uniquely written as $a_1x_1 + \cdots + a_nx_n$ for some a_i in F, the map $f : V \to W$, which is

$$f(a_1x_1 + \cdots + a_nx_n) = a_1y_1 + \cdots + a_ny_n$$

for a_i in F, is well defined.

Thus f is a homomorphism. Since $f(a_1x_1 + \cdots + a_nx_n)$ implies $a_1y_1 + \cdots + a_ny_n = 0$ implies $a_1 = \cdots = a_n = 0$, which in turn implies $a_1x_1 + \cdots + a_nx_n = 0$, therefore f is one to one. Then, since all elements of W is of the form $a_1y_1 + \cdots + a_ny_n$, which is equal to $f(a_1x_1 + \cdots + a_nx_n)$ for some $a_1, \ldots, a_n$ in F, therefore f is also onto. Hence f is an isomorphism. ¶

Definition 63 defines polynomial codes, Definition 56 matrix code, and Definition 65 parity check matrix.

Definition 63. Let $g(x) = g_0 + \cdots + g_kx^k$ be a polynomial in $F[x]$. We call the *polynomial code* with encoding or generating polynomial $g(x)$ a code which encodes each original word of the message $a = (a_0, \ldots, a_{b-1})$, corresponding to

$$a(x) = a_0 + \cdots + a_{b-1}x^{b-1}$$

into the code word $b = (b_0, \ldots, b_{b+k-1})$, which corresponds to the code polynomial

$$b(x) = b_0 + \cdots + b_{b+k-1}x^{b+k-1} = a(x)g(x)$$

§

Note 7. We assume for our generating polynomial that $g_0 \neq 0$ and $g_k \neq 0$. To justify this assumption, suppose we have $g(x) = g_0 + \cdots + g_kx^k$. If $g_0 = 0$, then we choose a new polynomial for $g(x)$ as $g_1(x) = a_1 + \cdots + a_kx^{k-1}$. If $g_k = 0$, then we choose another polynomial $g_2(x) = g_0 + \cdots + a_{k-1}x^{k-1}$. In either case our choice becomes more economical.

§

Theorem 37. A polynomial with coefficients in $\mathbf{B}$ is divisible by $1+x$ if and only if it has an even number of terms.

Proof. Let $f(x) = a_0 + \cdots + a_n x^n$ for all a_i in $\mathbf{B}$, $i = 1, \ldots, n$, and let $1 + x | f(x)$. Then there exists a polynomial $b(x)$ in $\mathbf{B}$ such that $f(x) \equiv (1+x)b(x)$. If $x = 1$, we have $a_0 + \cdots + a_n = 0$. Since the field $\mathbf{B}$ is of characteristic 2, this is only possible if the number of non-zero terms is even. Conversely, let $f(x)$ have an even number of non-zero terms, say $f(x) = x^{i_1} + \cdots + x^{i_{2k}}$, where $i_1 < \cdots < i_{2k}$. Rewrite this as

$$f(x) = \left(x^{i_1} + x^{i_2}\right) + \cdots + \left(x^{i_{2k-1}} + x^{i_{2k}}\right)$$

For $i < j$, $x^i + x^j = x^i\left(1 + x^{j-i}\right) = x^i(1+x)\left(1 + \cdots + x^{j-i-1}\right)$, which means that $1 + x | x^i + x^j$. Therefore $1+x$ divides all bracketed terms in $f(x)$, and hence $1 + x | f(x)$. ¶

Theorem 38. If $g(x) \in \mathbf{B}[x]$ divides no polynomials of the form $x^k - 1$ for $k < n$, then the binary polynomial code of length n generated by $g(x)$ has the minimum distance of at least 3.

Proof. Let $g(x) = g_0 + \cdots + g_r x^r$, where g_i are in $\mathbf{B}$, $g_0 \neq 0$ and $g_r \neq 0$. Let $b = n - r$. Suppose the opposite to what the theorem says is true. Then, polynomial code being a group code, there exists $b(x)$ with at most two non-zero entries. There are two cases to consider, namely $b(x) = x^i + x^j$, where $i < j$, and $b(x) = x^i$, where $i < n$. In the first one of these, since n is the code length, we have $j < n$, hence $0 < j - i < n$. Since $g(x) | b(x)$ implies $g(x) | x^j\left(1 + x^{j-i}\right)$, and $g_0 \neq 0$ implies $x \not| g(x)$, therefore $g(x) | 1 + x^{j-i}$ which contradicts our hypothesis. In the second case, similarly to the above $g(x) | x^i$ and we again have a contradiction. ¶

Definition 64. Let C be a (b, n)-code. If there exists a $b \times n$ matrix G of rank b such that $C = \left\{\mathbf{a}G | a \in \mathbf{B}^b\right\}$, then G is called a *generator matrix* of the code C, and C is called a *matrix code* generated by G.

§

Definition 65. Let C be a (b, n)-code. If there exists an $(n-b) \times n$ matrix H of rank $n - b$ such that $H\mathbf{b}^9 0 = 0$ for all $\mathbf{b}$ in C, then H is called a *parity check matrix* of C.

§

Theorem 39. A polynomial code is a matrix code.

Proof. Let C be a polynomial b, n-code with the encoding polynomial $g(x) = g_0 + \cdots + g_k x^k$. Then $n = b + k$. Let G be the $b \times n$ matrix whose first row begins with entries $g_0, \ldots, g_k$ followed by b zeros, and whose succeeding row is an anticlockwise cyclic shift of the previous one, that is

$$G = \begin{bmatrix} g_0 & g_1 & \cdots & g_k & 0 & \cdots & 0 \\ 0 & g_0 & & \cdots & g_k & & \\ \vdots & & & & & & \\ 0 & & \cdots & & g_0 & \cdots & g_k \end{bmatrix}$$

The determinant of the submatrix formed by the first b columns is non-zero, since $g_0 \neq 0$ and hence $g_0^b \neq 0$. Thus the rank of G is m. Let the original word to be coded be $a = (a_0, \ldots, a_{m-1})$. Then, since the code word generated by aG is the same as that generated from $a(x)g(x)$, the two codes are identical. ¶

Then we look at the Hamming codes as described in Algorithm 4.

Algorithm 4 gives a procedure that construct the Hamming codes. Here b_i represents a binary code equivalent of the decimal number i.

Algorithm 4 *Hamming codes*

choose r a positive integer
$b \leftarrow 2^r - r - 1$
$n \leftarrow 2^r - 1$
for $i = 1$ to $2^r - 1$ **do**
  (the i^{th} row of M)$\leftarrow (\mathbf{b}_i)$
endfor
for $i = 1$ to $2^r - 1$ **do**
  $(a_1, \ldots, a_{2^r-1}) \leftarrow (\mathbf{b}_i)$
  $(b_{2^{2-1}}, \ldots, b_{2^{r-2}} - 1, b_{2^{r-2}} + 1, \ldots, b_{2^{r-1}} - 1) \leftarrow (a_1, \ldots, a_{2^r-1})$
$(b_{2^{j-1}}; j = 1, \ldots, r) \leftarrow$ **solve** $(\mathbf{b}M = 0)$
the i^{th} code word $\leftarrow (b_1, \ldots, b_n)$
endfor

Note 8. Each code word in a Hamming code contains $b - n = 2^r - r - 1 - 2^r + 1 = r$ check digits. The value of r is called the *redundancy* of the code. §

Bibliography

L R Vermani. *Elements of algebraic codng theory.* Chapman & Hall, 1996

Examples
Group, polynomial, and Hamming codes

14^{th} January, 2007

31. Construct a Hamming code with three check digits.

Solution. Choose $r = 3$. Then the code words will have $n = 2^r - 1 = 2^3 = 7$ digits, and the message words $m = 2^r - r - 1 = 2^3 - 3 - 1 = 4$ digits. In each code word there are r check digits. The redundancy of the code is r. The check digits are formed as follows.

$$\overbrace{\begin{array}{cccccc} b_1 & b_2 & b_3 & b_4 & \cdots & b_n \\ \downarrow & \downarrow & & \downarrow & \downarrow & \\ b_{2^0} & b_{2^1} & & b_{2^2} & b_{2^{r-1}} & \end{array}}^{b}$$

$$\underbrace{\qquad\qquad\qquad\qquad\qquad}_{r \text{ check digits}}$$

The rest of the code word are the $2^r - r - 1$ message digits in their usual order. Then for our present problem.

$$\begin{array}{lccccccc} \text{check digits} & b_1 & b_2 & & b_4 & & & \\ & \uparrow & \uparrow & & \uparrow & & & \\ \text{code word} & b_1 & b_2 & b_3 & b_4 & b_5 & b_6 & b_7 \\ & & & \downarrow & & \downarrow & \downarrow & \downarrow \\ \text{message word} & & & a_1 & & a_2 & a_3 & a_4 \end{array}$$

Next, form a $(2^r - 1) \times r$ matrix M, where the i^{th} row is the binary representation of the number i.

$$M = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

Then form the matrix equation $\mathbf{b}M = 0$ which gives r linear equations in the r unknowns $b_1, b_2, \ldots, b_{2^{r-1}}$.

$$\begin{pmatrix} b_1 & b_2 & b_3 & b_4 & b_5 & b_6 & b_7 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} = \mathbf{0}$$

This gives us

$$b_4 + b_5 + b_6 + b_7 = 0$$
$$b_2 + b_3 + b_6 + b_7 = 0$$
$$b_1 + b_3 + b_5 + b_7 = 0$$

To encode a message word, we place the message in its proper positions, then find b_{2^i}, where $0 \leq i \leq r-1$. For example, the message $a_1\, a_2\, a_3\, a_4 = 1001$ yields

$$(\, b_1 \quad b_2 \quad 1 \quad b_4 \quad 0 \quad 0 \quad 1\,) \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} = \mathbf{0}$$

Then,

$$\begin{aligned} b_4 + 1 = 0 \quad &\rightarrow \quad b_4 = 1 \\ b_2 + 1 + 1 = 0 \quad &\rightarrow \quad b_2 = 0 \\ b_1 + 1 + 1 = 0 \quad &\rightarrow \quad b_1 = 0 \end{aligned}$$

Therefore the encoded message is 0011001.

#

Exercises for Group, polynomial, and Hamming codes

14^{th} January, 2007

32. Given a generator matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

Find the parity check matrix, then all code words generated by the generator matrix above, and then find the dual code to that code.

33. Check whether 974–93880–3–8, 974–93880–4–6 and 0–1392–4101–4 are ISBN's.

34. Find the missing digits in the following ISBN's, 974–912□7–4–3 and 974–9□3□1–8–4.

35. Construct tables of multiplicative inverses for $GF(5)$ and $GF(11)$.

36. Find a primitive element for each of $GF(5)$ and $GF(13)$.

37. Find the parity check matrix of the matrix code given by the generator matrix

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

38. The parity check matrix

$$H = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

defines a code $E : \mathbf{F}_2^3 \to \mathbf{F}_2^6$. Find all the code words. Does the code thus resulted correct all single errors?

39. The *exponent* of a polynomial $g(x) \in \mathbf{F}_2[x]$ is the least positive integer e such that $g(x)|x^e - 1$. Find the exponent of the polynomials $a(x)$, $b(x)$ and $c(x)$ in $\mathbf{B}[x]$, when $a(x) = 1 + x + x^2$, $b(x) = x + x^3$ and $c(x) = 1 + x^2 + x^4$.

40. Given an encoding polynomial $g(x) = 1 + x^2 + x^3$. Find the generator matrix G of the $(4, 7)$ polynomial code. Then find the parity check matrix H of the code thus generated.

41. Find the binary equivalence of the decimal numbers 543, 25, 87 and 166.

42. Find the decimal representation of the binary numbers 11001110, 10100101, 10001100 and 10111.

43. Find all the code words of the binary $(3, 5)$ Hamming code.

44. Find a parity check matrix of the binary $(9, 13)$ Hamming code.

45. Write a parity check matrix for the 7-ary $(8, 6)$-Hamming code, then use it to decode the received vectors 54326010 and 11063452.

Reference

Raymond Hill. *A first course in coding theory.* Clarendon, 1986

L R Vermani. *Elements of algebraic codng theory.* Chapman & Hall, 1996

Finite field- and BCH codes

2^{nd} December 2005

Definition 66. Let G be a group. Then a *coset* is a subgroup H of G which is either a *left coset* of H, that is $xH = \{xh : h \in H\}$ for some x in G, or a *right coset* $Hx = \{hx : h \in H\}$ of the same.

§

Definition 67. Let polynomials $f_1(x), \ldots, f_r(x)$ in $\mathbf{F}_q[x]$ be non-zero. Then the *least common multiple* $\operatorname{lcm}(f_1(x), \ldots, f_r(x))$ of $f_1(x), \ldots, f_r(x)$ is the monic polynomial of the lowest degree which is a multiple of all $f_i(x)$, $i = 1, \ldots, r$.

§

Problem 6. Prove that for non-zero polynomials $f_1(x), \ldots, f_r(x)$ in $\mathbf{F}_q[x]$,

$$\operatorname{lcm}(f_1(x), \ldots, f_r(x)) = \operatorname{lcm}(\operatorname{lcm}(f_1(x), \ldots, f_{r-1}(x)), f_r(x))$$

§

Note 9. Let $f_1(x), \ldots, f_r(x)$ in $\mathbf{F}_q[x]$ have the factorisations,

$$\begin{aligned} f_1(x) &= a_1 (p_1(x))^{e_{11}} \cdots (p_n(x))^{e_{1n}} \\ &\vdots \\ f_r(x) &= a_r (p_1(x))^{e_{r1}} \cdots (p_n(x))^{e_{rn}} \end{aligned}$$

where $a_1, \ldots, a_r$ are in $\mathbf{F}_q^*$, $e_{ij} \geq 0$, and $p_i(x)$ are distinct monic irreducible polynomials over $\mathbf{F}_q$, then

$$\operatorname{lcm}(f_1(x), \ldots, f_r(x)) = (p_1(x))^{\max(e_{11}, \ldots, e_{r1})} \cdots (p_n(x))^{\max(e_{1n}, \ldots, e_{rn})}$$

§

Theorem 40. Let $f(x), f_1(x), \ldots, f_r(x)$ be polynomials over $\mathbf{F}_q$. If $f(x)$ is divisible by every polynomial f_i, for $i = 1, \ldots, r$, then $f(x)$ is also divisible by $\operatorname{lcm}(f_1(x), \ldots, f_r(x))$.

Proof. Consider first the case where there are only two different polynomials, $f_1(x)$ and $f_2(x)$. The prime components of $f_1(x)$ and $f_2(x)$ may be grouped into those which are unique among them and those which are shared. Since $f(x) = u_1(x)f_1(x) + r_1(x)$ and $f(x) = u_2(x)f_2(x) + r_2(x)$, it follows that $f(x)$ contains both of these two groups of primes. In other words, $f(x) = u(x)\operatorname{lcm}(f_1(x), f_2(x)) + r(x)$.
Next, consider the case where there are more than two f_i's. Suppose for $f(x)$, that $f(x) = u_r(x)\operatorname{lcm}(f_1(x), \ldots, f_r(x))$. Then if we let

$$f_c(x) = \operatorname{lcm}(f_1(x), \ldots, f_r(x))$$

and if we introduce another polynomial $f_{r+1}(x)$ such that $f(x) = u_{r+1}f_{r+1} + r_{r+1}(x)$, then following the same line of reasoning as the above we have,

$$\operatorname{lcm}(f_1(x), \ldots, f_{r+1}(x)) | f(x)$$

¶

Definition 68. A non-empty subset S of a ring R is called a *subring* of R if the elements of S form a ring with respect to the operations defined in R.

§

Theorem 41. Let R be a ring. Then a non-empty subset S of R is a subring if and only if S is closed under addition, multiplication, and the formation of additive inverse.

Proof. Since S is a subset of R, additive associativity, identity and commutativity are inherited to S from R. The existence of the inverse for each element s in S is certain provided that the formation of an additive inverse is guaranteed. And similarly in the case of multiplication, both associativeness and distributiveness hold once we know that S is closed under multiplication.

¶

Definition 69. Let R be a ring. We call an *ideal* in R a subring I having such property that for all i in I, then both xi and ix are also in I for every element x in R. Further, if I is a proper subset of R, then it is called a *proper ideal*. By *trivial ideal* one means either the *zero ideal* $\{0\}$ consisting of the zero element alone, or the ring R itself.

§

Note 10. The significance of the ideals in a ring is that they let us construct other rings from the first. The cosets of a ring R is a partition of R into equivalence sets, which are non-empty and disjoint, the union of which is the whole of the ring R.

§

Definition 70. Let R be a ring and I an ideal in it. Then two elements x and y in R are said to be *congruent modulo* I, denoted by $x \equiv y (\mathrm{mod}\, I)$, if $x - y$ is in I. Since there is only ideal, we may a write this congruence as simply $x \equiv y$.

§

Note 11. The congruence modulo I of a ring R as defined in Definition 70 is an equivalence relation since it is true that $x \equiv x$ for every x; $x \equiv y$ implies $y \equiv x$; and $x \equiv y$ and $y \equiv z$ implies $x \equiv z$.

§

Note 12. Congruences can be added and multiplied as if they were ordinary equations. In other words, if $x_1 \equiv x_2$ and $y_1 \equiv y_2$, then $x_1 + y_1 \equiv x_2 + y_2$ and $x_1 y_1 \equiv x_2 y_2$.

§

Definition 71. Let R be a ring and let x be an element of R. Then the

coset $[x]$ containing x is the set of all elements y such that $y \equiv x$. Then,

$$\begin{aligned}[x] = \{y : y \equiv x\} &= \{y : y - x \in I\} \\ &= \{y : y - x = i \text{ for some } i \in I\} \\ &= \{y : y = x + i \text{ for some } i \in I\} \\ &= \{x + i : i \in I\} = x + I\end{aligned}$$

Furthermore, $[x] = [x_1]$ means that $x \equiv x_1$, that is to say, $x - x_1$ is in I. Here x and x_1 are called *representatives* of the coset which contains them.

§

Definition 72. A *quotient ring*, aka *residue-class-*, *factor-*, or *difference ring*, is a ring having the form of a quotient A/i of a ring A and one of its ideal i. In other words, the quotient ring of R with respect to I the ring $R/I = \{x + I : x \in R\}$, where $x+I = \{x + i : i \in I\}$ is the coset of an element x in R, and where addition and multiplication are defined as,

$$[x] + [y] = [x + y]$$

and

$$[x] \cdot [y] = [xy]$$

§

Theorem 42. The zero element of R/I is $0 + I = I$, the negative of $x + I$ is $(-x) + I$. If R is commutative, then R/I is also commutative. If R has an identity 1 and a proper ideal I, then R/I has an identity $1 + I$.

§

Problem 7. Prove Theorem 42.

§

Theorem 43. Let R be a ring and I an ideal of R. Then, for x and y in R,

$$(x + I) + (y + I) = (x + y) + I$$

and

$$(x + I)(y + I) = xy + I$$

Proof. Let a and b be any two elements of the ideal I. Then,

$$(x + a) + (y + b) = x + a + y + b = (x + y) + (a + b) = (x + y) + p$$

where $p = a + b$ is in I. Further,

$$\begin{aligned}(x + a)(y + b) &= xy + bx + ay + ab \\ &= xy + c + d + e = xy + f\end{aligned}$$

where $c = bx$, $d = ay$, $e = ab$ and $f = c + d + e$ are all elements of I. ¶

Note 13. Theorem 43 and Note 12 show that the quotient ring R/I defined in Definitions 72 is independent of the choice of x and y in the cosets $x+I$ and $y+I$. In other words, the cosets $[x+y]$ and $[xy]$ resulted from addition and respectively multiplication in no ways depend on the particular representatives x and y chosen for the cosets $[x]$ and $[y]$ that go into them. This means that, if $x_1 \equiv x$ and $y_1 \equiv y$, then $[x_1+y_1]=[x+y]$ and $[x_1y_1]=[xy]$, or equivalently $x_1+y_1 \equiv x+y$ and $x_1y_1 \equiv xy$.

§

Example 31. Some examples of quotient ring are $\mathbf{Z}_2 = \mathbf{Z}/2\mathbf{Z}$ and $\mathbf{Z}_6 = \mathbf{Z}/6\mathbf{Z}$.

Theorem 44. The polynomial ring $F[x]$ is a commutative ring with identity.

Proof. $F[x]$ is a ring over the field F since under addition it is closed, associative and commutative, and has 0 as the identity and the inverse $-f(x)$, where $f(x) \in F[x]$; and under multiplication it is associative, distributive and commutative, and has 1 as the identity. ¶

Definition 73. Let R be a commutative ring with identity. Then for any a in R the *principal ideal* generated by a is $\langle a \rangle = aR = \{ar : r \in R\}$. Further, R is called *principal ideal ring* if all its ideals are of this form.

§

Theorem 45. Let F be a field. Then the polynomial ring $F[x]$ is a principal ideal ring.

Proof. The polynomial ring $F[x]$ being a commutative ring with identity, it remains only to show that all its ideals are of the form $\langle a \rangle R = aR = \{ar : r \in R\}$, where a is in R. Let I be an ideal of $F[x]$. If $I=0$, then I is a principal ideal generated by 0. If $I \neq 0$, then choose $0 \neq f(x) \in I$ such that $\deg f \leq \deg g$ for all non-zero $g(x)$ in I. Write $g(x) = q(x)f(x) + r(x)$. If $\deg g < \deg f$, then $q=0$ and $r=f$. On the other hand, if $n = \deg f \leq \deg g$, then either r is 0 or $\deg r < \deg f$. Let

$$f(x) = a_0x^n + \cdots + a_n$$

and

$$g(x) = b_0x^m + \cdots + b_m$$

Then, with $a_0 \neq 0$,

$$g(x) = a_0^{-1}b_0x^{m-n}f(x) + g_1(x) \tag{23}$$

where $\deg g_1 \leq m-1$. Then

$$g_1(x) = q_1(x)f(x) + r(x) \tag{24}$$

From this it follows that either $r=0$ or $\deg r < \deg f$. From Equation's 23 and 24, $g(x) = q(x)f(x) + r(x)$, where $q(x) = a_0^{-1}b_0x^{m-n} + q_1$ is in $F[x]$. If $r \neq 0$, then $r(x)$ is in I and $\deg r < \deg f$, which contradicts our choice of $f(x)$. Therefore $g = qf$ and I is a principal ideal generated by $f(x)$. ¶

Definition 74. Let R be a commutative ring with identity. Then a non-constant $f(x)$ in $R[x]$ is said to be *reducible* if, for some $g(x)$ and $h(x)$ in $R[x]$, $f(x) = g(x)h(x)$ implies either $\deg g(x) = 0$ or $\deg h(x) = 0$. Otherwise $f(x)$ is said to be *reducible*.

§

Theorem 46. Let F be a field $f(x)$ in $F[x]$ an irreducible polynomial. Then $F[x]/\langle f(x)\rangle$ is a field.

Proof. Let I be the ideal $\langle f(x)\rangle$ of $F[x]$ generated by $f(x)$. If $I = F[x]$, then $f(x)$ has an inverse, that is $1 = f(x)g(x)$ for some $g(x)$ in $F[x]$. Then $f(x)$ is a constant polynomial, which contradicts our statement of the theorem. Therefore $F[x]/I$ has at least two elements, and $F[x]/I$ being a polynomial ring it is a commutative ring with identity. Let $g \in F[x]$ and $g \notin I$. Then,

$$J = \{a(x)f(x) + b(x)g(x) : a(x), b(x) \in F[x]\}$$

is an ideal of $F[x]$ and there exists $h(x)$ in $F[x]$ such that $J = \langle h(x)\rangle$. But $f(x) = 1f(x) + 0g(x)$ is in J, and thus $f(x) = a(x)h(x)$ for some $a(x)$ in $F[x]$. The polynomial $f(x)$ being irreducible, either $\deg h(x) = 0$ or $\deg a(x) = 0$. If the latter is the case, then $a(x)$ is a unit in $F[x]$, and then $h(x)$ is in I, hence $J = I$, and hence a contradiction since we began with g being in J but not in I. Therefore it must be the case that $h(x)$ is a unit in $F[x]$, hence J is a unit, and thus $1 = a(x)f(x) + b(x)g(x)$ for some $a(x)$ and $b(x)$ in $F[x]$. And then $1 + I = I + b(x)g(x) = (b(x) + I)(g(x) + I)$. Thus $g(x) + I$ has an inverse and $F[x]/I$ is a field. ¶

Definition 75. Let K be a field and F a subfield of K. Then K is called an *extension* of the field F, denoted by $K|_F$. Since K has multiplication, it is a vector space over F. The dimension of the vector space K over F is called the *degree* $[K : F]$ of the extension K of F. The extension $K|_F$ is said to be *finite* if the degree $[K : F]$ is finite.

§

Definition 76. A *prime subfield* of a field F is the intersection of all subfields of F. It is the smallest of all subfields of F, and is unique. A *prime field* is a field which has no proper subfields.

§

Definition 77. Let $K|_F$ be an extension of a field F. Then $\alpha \in K$ is said to be *algebraic* over F if there exists $f(x)$ in $F[x]$ which has α as a root. Let α in K be algebraic over F and consider $A = \{f(x) \in F[x] : f(\alpha) = 0\}$. Here A is an ideal of the principal ideal domain $F[x]$. Let $m_1(x)$ in $F[x]$ be a generator of A. If a is the coefficient of the highest power of x in $m_1(x)$, then $m(x) = a^{-1}m_1(x)$ is a monic polynomial with $\deg m(x) = \deg m_1(x)$, and $m(x)$ is also a generator of A. Let $m(x) = r(x)s(x)$ for some $r(x)$ and $s(x)$ in $F[x]$. Then either $r(\alpha) = 0$ or $s(\alpha) = 0$, that is either $m(x)|r(x)$ or $m(x)|s(x)$. But $\deg m = \deg r + \deg s$, therefore either $\deg r(x) = 0$ or

$\deg s(x) = 0$. Hence $m(x)$ is irreducible. Since $m(x)$ is monic, irreducible and is of the least degree possible while admitting α as a root, therefore $m(x)$ is called the *minimal polynomial* of α over $F[x]$.

§

Theorem 47. Let C be an (n, k) linear code over F_q with parity-check matrix H, and $d(C)$ the smallest number of column of H that are linearly dependent. Then if every subset of $2t$ or fewer columns of H is linearly independent, the code is capable of correcting all error patterns of weight $w \leq t$.

Proof. When $q = 2$, linear independence amounts to summing to $\mathbf{0}$. The code words of C are those vectors $\mathbf{x}$ in $V_n(F_q)$ for which $H\mathbf{x}^T = \mathbf{0}$. But $H\mathbf{x}^T$ is a linear combination of the columns of H, that is to say, if $H = [\mathbf{c}_1 \quad \cdots \quad \mathbf{c}_n]$, then $H\mathbf{x}^T = x_1\mathbf{c}_1 + \cdots + x_n\mathbf{c}_n$. Hence a non-zero code word of weight w gives a nontrivial linear dependence among w columns of H, and vice versa. ¶

Corollary 47[1]. If $q = 2$ and all possible linear combinations of up to e columns are distinct, then $d(C) \geq 2e + 1$, and C can then correct all patterns of weight e or less.

Problem 8. Prove Corollary 47[1].

§

Note 14. Hamming codes correct single errors. An extension of this is to the Bose-Chaudhuri-Hocquenghem codes which could correct multiple errors. In the case of Hamming code of length $n = 2^m - 1$, the parity-check matrix is given by $H = [\mathbf{v}_0 \quad \ldots \quad \mathbf{v}_{n-1}]$, where $(\mathbf{v}_0 \quad \cdots \quad \mathbf{v}_{n-1})$ is some ordering of the $2^m - 1$ non-zero column vectors in $V_m = V_m(F_2)$. The $m \times n$ matrix H takes m parity-check bits for the code to be able to correct one error. We may extend H such that it has m more rows and could correct two errors. Then,

$$H_2 = \begin{bmatrix} \mathbf{v}_0 & \cdots & \mathbf{v}_{n-1} \\ \mathbf{w}_0 & \cdots & \mathbf{w}_{n-1} \end{bmatrix}$$

where $\mathbf{w}_0, \ldots, \mathbf{w}_{n-1}$ are in V_m. Since $\mathbf{v}_i$'s are distinct, we may look at the mapping from $\mathbf{v}_i$ to $\mathbf{w}_i$ as a function from V_m into itself, then

$$H_2 = \begin{bmatrix} \mathbf{v}_0 & \cdots & \mathbf{v}_{n-1} \\ \mathbf{f}(\mathbf{v}_0) & \cdots & \mathbf{f}(\mathbf{v}_{n-1}) \end{bmatrix}$$

Then H_2 will define a code which corrects two errors if and only if the syndromes of the $1 + n + \binom{n}{2}$ error patterns of weights 0, 1 and 2 are all distinct. Any such syndrome is a sum of a subset of columns of H_2, and therefore a vector in V_{2m}. Let the syndrome be $\mathbf{s} = (s_1 \quad \ldots \quad s_{2m}) = (\mathbf{s}_1 \quad \mathbf{s}_2)$, where $\mathbf{s}_1 = (s_1, \ldots, s_m)$ and $\mathbf{s}_2 = (s_{m+1}, \ldots, s_{2m})$ are both in V_m. Defining

$\mathbf{f}(\mathbf{0},\mathbf{0}) = \mathbf{0}$ we consider a pair of errors occuring at i^{th}- and j^{th} position's, $\mathbf{s} = (\mathbf{v}_i + \mathbf{v}_j, \mathbf{f}(\mathbf{v}_i) + \mathbf{f}(\mathbf{v}_j))$. Then the system of equations,

$$\begin{aligned} \mathbf{u} + \mathbf{v} &= \mathbf{s}_1 \\ \mathbf{f}(\mathbf{u}) + \mathbf{f}(\mathbf{v}) &= \mathbf{s}_2 \end{aligned}$$

has at most one solution $(\mathbf{u}, \mathbf{v})$ for each pair of vectors from V_m. By trial and error we may find neither the linear mapping $\mathbf{f}(\mathbf{v}) = T\mathbf{v}$ nor the nonlinear polynomial of degree 2 works, but $\mathbf{f}(\mathbf{v}) = \mathbf{v}^3$ does. The matrix

$$H_2 = \begin{bmatrix} \alpha_0 & \cdots & \alpha_{n-1} \\ \alpha_0^3 & \cdots & \alpha_{n-1}^3 \end{bmatrix}$$

is the parity-check matrix of a binary code of length $n = 2^m - 1$ which corrects up to two errors. A vector $\mathbf{c} = (\, c_0 \quad \cdots \quad c_{n-1}\,)$ in $V_n(F_2)$ is a code word in the code defined by H_2 if and only if $\sum_{i=0}^{n} c_i\alpha_i = \sum_{i=0}^{n} c_i\alpha_i^3 = 0$. Since the $2m$ rows of the matrix H_2 over F_2 may not be all linearly independent, the dimension of the code is $d(C) \geq n - 2m = 2^m - 1 - 2m$.

§

Definition 78. The *Vandermonde matrix* is defined as

$$A = \begin{bmatrix} 1 & \cdots & 1 \\ a_1 & \cdots & a_r \\ \vdots & \ddots & \vdots \\ a_1^{r-1} & \cdots & a_r^{r-1} \end{bmatrix}$$

§

Theorem 48. Let $a_1, \ldots, a_r$ be distinct non-zero elements of a field. Then the the Vandermonde matrix is such that

$$\begin{vmatrix} [1 & \cdots & 1 \\ a_1 & \cdots & a_r \\ \vdots & \ddots & \vdots \\ a_1^{r-1} & \cdots & a_r^{r-1} \end{vmatrix} \neq 0$$

Proof. Subtracting $\text{row}(i+1) - a_1 \,\text{row}\, i$, $i = 1, \cdots, r-1$, yields,

$$\begin{aligned}
\det A &= \begin{vmatrix} 1 & 1 & \cdots & 1 \\ 0 & a_2 - a_1 & \cdots & a_r - a_1 \\ 0 & a_2(a_2 - a_1) & \cdots & a_r(a_r - a_1) \\ \vdots & & \ddots & \vdots \\ 0 & a_2^{r-2}(a_2 - a_1) & \cdots & a_r^{r-2}(a_r - a_1) \end{vmatrix} \\
&= (a_2 - a_1) \cdots (a_r - a_1) \begin{vmatrix} 1 & \cdots & 1 \\ a_2 & \cdots & a_r \\ \vdots & \ddots & \vdots \\ a_2^{r-2} & \cdots & a_r^{r-2} \end{vmatrix} \\
&= (a_2 - a_1) \cdots (a_r - a_1) \cdot (a_3 - a_2) \cdots (a_r - a_2) \begin{vmatrix} 1 & \cdots & 1 \\ a_3 & \cdots & a_r \\ \vdots & \ddots & \vdots \\ a_3^{r-3} & \cdots & a_r^{r-3} \end{vmatrix} \\
&\;\;\vdots \\
&= \prod_{i>j} (a_i - a_j)
\end{aligned}$$

Then, since a_i are distinct and non-zero, therefore $\det A$ is non-zero. ¶

Theorem 49. Any square matrix having a non-zero determinant has all its columns linearly independent.

Proof. Let A be an $r \times r$ matrix, and that $|A| \neq 0$. Then suppose the columns of A are linearly dependent. Then one may write some column of A as a linear combination of the others, for example

$$\mathbf{c}_j = \sum_{\substack{i=1 \\ i \neq j}}^{r} a_i \mathbf{c}_i$$

Then if column $\mathbf{c}_j$ is replaced by $\mathbf{c}_j - \sum_{\substack{i=1 \\ i \neq j}}^{r} a_i \mathbf{c}_i$ gives a matrix B with $|B| = |A|$. But B also has a column whose all elements are zeros, which means that $|A| = |B| = 0$, a contradiction and thus the proof. ¶

Theorem 50. Let $(\alpha_0, \ldots, \alpha_{n-1})$ be an ordering of non-zero elements of $\mathbf{F}_{2^m}$, and let t be a positive integer such that $t \leq 2^{m-1} - 1$. Then the matrix

$$H = \begin{bmatrix} \alpha_0 & \cdots & \alpha_{n-1} \\ \alpha_0^3 & \cdots & \alpha_{n-1}^3 \\ \vdots & \ddots & \vdots \\ \alpha_0^{2t-1} & \cdots & \alpha_{n-1}^{2t-1} \end{bmatrix}$$

is the parity-check matrix of a binary (n, k)-code capable of correcting all error patterns of weight $w \leq t$, with dimension $k \geq n - mt$.

Proof. A vector $\mathbf{c} = (c_0, \ldots, c_{n-1})$ in $V_n(F_2)$ is a code word if and only if $H\mathbf{c}^T = \mathbf{0}$. Thus,

$$\sum_{i=0}^{n-1} c_i \alpha_i^j = 0$$

for $j = 1, 3, \ldots, 2t-1$. We simplify this by using the fact that $(x+y)^2 = x^2+y^2$ in characteristic 2, and $x^2 = x$ in F_2. Hence,

$$\left(\sum_{i=0}^{n-1} c_i \alpha_i^j \right)^2 = \sum_{i=0}^{n-1} c_i^2 \alpha_i^{2j} = \sum_{i=0}^{n-1} c_i \alpha_i^{2j}$$

for $j = 1, 3, \ldots, 2t-1$, which gives us

$$\sum_{i=0}^{n-1} c_i \alpha_i^j$$

for $j = 1, 2, \ldots, 2t$. Therefore we could also use the parity-check matrix

$$H^90 = \begin{bmatrix} \alpha_0 & \cdots & \alpha_{n-1} \\ \alpha_0^2 & \cdots & \alpha_{n-1}^3 \\ \vdots & \ddots & \vdots \\ \alpha_0^{2t} & \cdots & \alpha_{n-1}^{2t} \end{bmatrix}$$

According to Theorem 47 H^90 is a parity-check matrix which corrects t errors if and only if every subset of $2t$ or fewer columns of H^90 is linearly independent. Next, since a subset of $r \leq 2t$ columns of H^90 has the form

$$A = \begin{bmatrix} a_1 & \cdots & a_r \\ a_1^2 & \cdots & a_r^2 \\ \vdots & \ddots & \vdots \\ a_1^{2t} & \cdots & a_r^{2t} \end{bmatrix}$$

where $a_1, \ldots, a_r$ are distinct non-zero elements of F_{2m}, we may consider the matrix

$$A^90 = \begin{bmatrix} a_1 & \cdots & a_r \\ \vdots & \ddots & \vdots \\ a_1^r & \cdots & a_r^r \end{bmatrix}$$

which is nonsingular since its determinant by the Vandermonde determinant theorem, Theorem 48, is

$$\det A^90 = a_1 \cdots a_r \begin{vmatrix} 1 & \cdots & 1 \\ a_1 & \cdots & a_r \\ \vdots & \ddots & \vdots \\ a_1^{r-1} & \cdots & a_r^{r-1} \end{vmatrix} = a_1 \cdots a_r \prod_{i<j} (a_j - a_i) \neq 0$$

Then the columns of A^90, and hence those of A, cannot be linearly dependent, and therefore the code corrects all error patterns of weight up to t. Now H, as a matrix with entries from F_2 rather than F_{2m}, has dimensions $mt \times n$, hence the dual code has dimension $k \leq mt$, and the code has dimension $k \geq n - mt$. ¶

Theorem 51. Let C be a linear (n, k)-code over $GF(q)$ with parity-check matrix H. Then the minimum distance of C is d if and only if any $d - 1$ columns of H are linearly independent but some d columns are linearly dependent.

Proof. The minimum distance of a code $d(C)$ is equal to the minimum of the weights of the non-zero code words. Let $\mathbf{x} = x_1 \cdots x_n$ be a vector in $V(n, q)$. Then $\mathbf{x}$ is in C if and only if $\mathbf{x}H^T = \mathbf{0}$ if and only if $x_1\mathbf{h}_1 + \cdots + x_n\mathbf{h}_n = \mathbf{0}$, where $\mathbf{h}_1, \ldots, \mathbf{h}_n$ are the columns of H. Therefore there is a set of d linearly dependent columns of H corresponding to each code word $\mathbf{x}$ of weight d. On the other hand, if there existed a set of $d - 1$ linearly dependent columns of H, then there would exist some scalars $x_{i_1}, \ldots, x_{i_{d-1}}$, not all zero, such that $\sum_{j=1}^{d-1} x_{i_j} = \mathbf{0}$. But if this were the case, then $\mathbf{x}H^T = \mathbf{0}$ and so would be a code word of weight $0 < d < d(C)$. ¶

Theorem 52. The maximum dictionary size m such that there exists a q-ary (n, m, d)-code is $A_q(n, d) \leq q^{n-d+1}$.

Proof. Let C be a q-ary (n, m, d)-code. If we remove the last $d-1$ coordinates from each code word, then the m vectors of length $n - d + 1$ so obtained must be distinct, otherwise $d(C)$ must be less than d, which would contradict the statement above. Therefore $m \leq q^{n-d+1}$. ¶

Theorem 53. Let C be the code over $GF(q)$, where q is a prime number, and C is defined to have the parity-check matrix

$$H = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 3 & \cdots & n \\ 1 & 2^2 & 3^2 & \cdots & n^2 \\ \vdots & & & \ddots & \vdots \\ 1 & 2^{d-2} & 3^{d-2} & \cdots & n^{d-2} \end{bmatrix}$$

where $d \leq n \leq q - 1$. If q is a prime-power, then $A_q(n, d) = q^{n-d+1}$.

Proof. We have,

$$C = \left\{ x_1 \cdots x_n \in V(n, q) \text{ s.t. } \sum_{i=1}^{n} i^j x_i = 0 \text{ for } j = 0, 1, \ldots, d - 2 \right\}$$

Any $d - 1$ columns form a Vandermonde matrix, and therefore by Theorem's 48 and 49 are linearly independent. By Theorem 51 C has a minimum distance d and therefore is a q-ary (n, q^{n-d+1}, d)-code. The proof follows since C meets the Singleton bound of Theorem 52. ¶

Problem 9. Find the decoding procedure for the BCH codes.

Solution. Assume that $d = 2t + 1$ and H has $2t$ rows. Suppose the code word $\mathbf{c} = c_1 \cdots c_n$ is transmitted and the vector $\mathbf{r} = r_1 \cdots r_n$ is received. Assuming that at most t errors have occurred, let $x_1, \ldots, x_t$ be their positions and $m_1, \ldots, m_t$ their respective magnitudes. Then the syndrome is

$$(s_1, \ldots, s_{2t}) = \mathbf{r}H^T$$

and we have

$$s_j = \sum_{i=1}^{n} r_i i^{j-1} = \sum_{i=1}^{t} m_i x_i^{j-1} \tag{25}$$

for $j = 1, \ldots, 2t$. Then from

$$\phi(\theta) = \frac{m_1}{1 - x_1\theta} + \frac{m_2}{1 - x_2\theta} + \cdots + \frac{m_t}{1 - x_t\theta} \tag{26}$$

and

$$\frac{m_i}{1 - x_i\theta} = m_i \left(1 + x_i\theta + x_i^2\theta^2 \cdots\right)$$

together with Equation 25, we have

$$\phi(\theta) = s_1 + s_2\theta + \cdots + s_{2t}\theta^{2t-1} + \cdots$$

Also, from Equation 26 we have

$$\phi(\theta) = \frac{a_1 + a_2\theta + a_3\theta^2 + \cdots + a_t\theta^{t-1}}{1 + b_1\theta + b_2\theta^2 + \cdots + b^t\theta^t} \tag{27}$$

Hence,

$$\left(s_1 + s_2\theta + s_3\theta^2 + \cdots\right)\left(1 + b_1\theta + b_2\theta^2 + \cdots + b_t\theta^t\right) = a_1 + a_2\theta + \cdots + a_t\theta^{t-1}$$

Which gives us

$$a_1 = s_1 \quad \text{and} \quad a_i = \sum_{j=0}^{i-1} s_{i-j} b_i, \quad i = 2, \ldots, t \tag{28}$$

and

$$0 = \sum_{j=0}^{t} s_{i-j} b_j, \quad i = t+1, \ldots, 2t \tag{29}$$

With a_i and b_i known we may turn Equation 27 into partial fractions

$$\phi(\theta) = \frac{p_1}{1 - q_1\theta} + \cdots + \frac{p_t}{1 - q_t\theta}$$

and therefore $m_i = p_i$ and $x_i = q_i$, for $i = 1, \ldots, t$, and the system in Equation 25 is solved. Algorithm 5 then gives the procedure for error correction.

#

Note 15. The polynomial

$$\sigma(\theta) = 1 + b_1\theta + b_2\theta^2 + \cdots + b_t\theta^t = (1 - x_1\theta)\cdots(1 - x_t\theta) \tag{30}$$

can be used to locate the location of the errors. The polynomial

$$\omega(\theta) = a_1 + a_2\theta + \cdots + a_t\theta^{t-1}$$

can be used to find the magnitude of the errors.

§

Algorithm 5 *Procedure for correcting up to t errors in BCH codes.*

input: **r**
find $s_1, \ldots, s_{2t}$
$e \leftarrow$maximum number of equations in Equation 29
for $i = e + 1$ to t **do**
  $b_i \leftarrow 0$
endfor
$(b_1, \ldots, b_e) \leftarrow$**solve** the first e equations of Equation 29
$(z_1, \ldots, z_e) \leftarrow$**find** the e zeros of Equation 30
$(a_1, \ldots, a_e) \leftarrow$**solve** Equation 28
for $i = 1$ to e **do**
  $m_i \leftarrow \frac{a_1 + a_2 x_i + \cdots + a_e x_i^{e-1}}{\prod_{\substack{j=1 \\ j \neq i}}^{e} (1 + x_j x_i)}$
endfor

Bibliography

Raymond Hill. *A first course in coding theory.* Clarendon, 1986

San Ling and Chaoping Xing. *Coding theory, a first course.* Cambridge University Press, 2004

Robert J McEliece *The theory of information and coding.* Addison-Wesley, 1977

George F Simmons. *Topology and modern analysis.* McGraw-Hill, 1963

L R Vermani. *Elements of algebraic codng theory.* Chapman & Hall, 1996

Examples
Finite field and BCH codes

14^{th} January, 2007

46. Construct a binary BCH code of length 7 and minimum distance 3.

Solution. Here $n = 7$ and, the code being binary, q=2. Choosing r smallest such that $q^r \geq n + 1$, we have $2^r \geq 7 + 1 = 8$ and $r = 3$. Suppose $x^3 + x + 1$ is reducible, then it must have either x or $x + 1$ as a factor, and then $x = 0$ or 1 would be its root. But $x|x^3 + x + 1$ gives a remainder 1 and so does $x + 1|x^3 + x + 1$. Thus neither of these divides $x^3 + x + 1$, therefore neither is a factor of the latter, hence $x^3 + x + 1$ is irreducible.

We have $p = 2$ and $n = 3$, hence $x^{p^n-1} = x^{8-1} - 1 = x^7 - 1$.

$$
\begin{array}{r|l}
x^3 + x + 1 & x^4 + x^2 + x + 1 \\
\hline
 & x^7 - 1 \\
 & x^7 + x^5 + x^4 \\
\hline
 & x^5 + x^4 + 1 \\
 & x^5 + x^3 + x^2 \\
\hline
 & x^4 + x^3 + x^2 + 1 \\
 & x^4 + x^2 + x \\
\hline
 & x^3 + x + 1 \\
 & 0
\end{array}
\quad \Rightarrow \quad x^3 + x + 1 | x^7 - 1
$$

For $k < 7$; if $k = 6$;

$$
\begin{array}{r|l}
x^3 + x + 1 & x^3 + x + 1 \\
\hline
 & x^6 - 1 \\
 & x^6 + x^4 + x^3 \\
\hline
 & x^4 + x^3 + 1 \\
 & x^4 + x^2 + x \\
\hline
 & x^3 + x^2 + x + 1 \\
 & x^3 + x + 1 \\
\hline
 & x^2
\end{array}
\quad \neq 0 \quad \Rightarrow \quad x^3 + x + 1 \not| x^6 - 1
$$

If $k = 5$;

$$
\begin{array}{r|l}
x^3 + x + 1 & x^2 + 1 \\
\hline
 & x^5 - 1 \\
 & x^5 + x^3 + x^2 \\
\hline
 & x^3 + x^2 + 1 \\
 & x^3 + x + 1 \\
\hline
 & x^2 + x
\end{array}
\quad \neq 0 \quad \Rightarrow x^3 + x + 1 \not| x^5 - 1
$$

If $k = 4$;

$$
\begin{array}{r|l}
x^3 + x + 1 & x \\
\hline
 & x^4 - 1 \\
 & x^4 + x + 1
\end{array}
\quad \neq 0 \quad \Rightarrow \quad x^3 + x + 1 \not| x^4 - 1
$$

When $k = 3$, $x^3 + x + 1 \not| x^3 - 1$ is obvious. Therefore $\alpha = x + \langle x^3 + x + 1 \rangle$ is a primitive. Then α satisfies $\alpha^3 + \alpha + 1 = 0$.

A minimum polynomial is a monic, irreducible polynomial of a least possible degree which has α as a root. For a finite field F or order p^n with k as its prime subfield, α and α^p have the same minimum polynomial over k for every $\alpha \in F$.

Since $p = 2$, therefore both α and α^2 have the same minimum polynomial. Then the generating polynomial is $x^3 + x + 1$. Let our message word be $a_0a_1a_2a_3$. Then the message polynomial is $a(x) = a_0 + a_1x + a_2x^2 + a_3x^3$, and the corresponding code polynomial $a(x)\left(x^3 + x + 1\right)$. Therefore the code word is

$$a_0 + (a_0 + a_1)\, x + (a_1 + a_2)\, x^2 + (a_0 + a_2 + a_3)\, x^3 + (a_1 + a_3)\, x^4 + a_2x^5 + a_3x^6$$

So our code word is

$$(a_0, (a_0 + a_1), (a_1 + a_2), (a_0 + a_2 + a_3), (a_1 + a_3), a_2, a_3)$$

Since the encoding polynomial has 3 non-zero terms, therefore the code has a minimum distance 3.

#

Linear codes

9^{th} December 2005

Definition 79. Let V be a vector space over $\mathbf{F}_q$. Then a set of vectors $A = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ in V is said to be *linearly independent* if and only if a *linear combination* $\lambda_1\mathbf{v}_1 + \cdots + \lambda_k\mathbf{v}_k$ being a zero-vector implies that λ_i, $i = 1, \ldots, k$, are zero.

§

Definition 80. Let V be a vector space over $\mathbf{F}_q$. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a non-empty subset of V. Then, the *linear span* $\langle S\rangle$ of S is defined as

$$\langle S\rangle = \left\{ \sum_{i=1}^{k} \lambda_i \mathbf{v}_i : \lambda_i \in \mathbf{F}_q \right\}$$

We say that the span $\langle S\rangle$ of S is a subset of V generated or spanned by S. Let C be a subspace of V, then a subset S of C is called a *generating-* or *spanning set* of C if $C = \langle S\rangle$.

§

Definition 81. An *inner product* on $\mathbf{F}_q$ is a mapping $\langle \mathbf{a}, \mathbf{b}\rangle : \mathbf{F}_q^n \times \mathbf{F}_q^n \to \mathbf{F}_q$ such that, for all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in $\mathbf{F}_q^n$,

a. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{u}, \mathbf{w}\rangle + \langle \mathbf{v}, \mathbf{w}\rangle$
b. $\langle \alpha\mathbf{v}, \mathbf{w}\rangle = \alpha \langle \mathbf{v}, \mathbf{w}\rangle$, where α is a scalar
c. $\langle \mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{w}, \mathbf{v}\rangle$
d. $\langle \mathbf{u}, \mathbf{v}\rangle = 0$, for all non-zero $\mathbf{u}$ in $\mathbf{F}_q^n$, if and only if $\mathbf{v} = \mathbf{0}$

§

Definition 82. Let $\mathbf{v}$ and $\mathbf{w}$ be two vectors in $\mathbf{F}_q^n$. Then the *scalar product*, aka the *dot-* or *Euclidean inner product*, between $\mathbf{v}$ and $\mathbf{w}$ is defined as $\mathbf{v} \cdot \mathbf{w} = \sum_{i=1}^{n} v_i w_i \in \mathbf{F}_q$. The two vectors are said to be *orthogonal* to each other if and only if $\mathbf{v} \cdot \mathbf{w} = 0$. The *orthogonal complement* $S^\perp$ of a non-empty subset S of $\mathbf{F}_q^n$, is defined to be

$$S^\perp = \left\{ \mathbf{v} \in \mathbf{F}_q^n : \mathbf{v} \cdot \mathbf{s} = 0 \,\text{for all}\, \mathbf{s} \in S \right\}$$

When $S = \emptyset$ we define $S^\perp = \mathbf{F}_q^n$.

§

Note 16. The orthogonal complement $S^\perp$ of a non-empty subset S of a vector space $\mathbf{F}_q^n$ is always a subspace of $\mathbf{F}_q^n$. Moreover, $\langle S\rangle^\perp = S^\perp$.

§

Definition 83. Let V be a vector space over $\mathbf{F}_q$. Then a non-empty subset $A = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of V is called a *basis* for V if $V = \langle A\rangle$ and A is linearly independent.

§

Theorem 54. Let V be a vector space over $\mathbf{F}_q$. If $\dim v = k$, then V has q^k elements and

$$\frac{1}{k!}\prod_{i=0}^{k-1}\left(q^k - q^i\right)$$

different bases.

Proof. If the basis for V is $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ and $\lambda_1, \ldots, \lambda_k$ are in $\mathbf{F}_q$, then $V = \sum_{i=1}^{k} \lambda_i \mathbf{v}_i$. Since $|\mathbf{F}_q|$ is q, there are q choices for each λ_i. Therefore V has exactly q^k elements.
Let $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis for V. Since B is non-empty, $\mathbf{v}_1 \neq \mathbf{0}$ and there are $q^k - 1$ choices for $\mathbf{v}_1$. Then there are $q^k - q^{i-1}$ choices of $\mathbf{v}_i$, for $i = 2, \ldots, k$ because $\mathbf{v}_i \notin \langle \mathbf{v}_1, \ldots, \mathbf{v}_{i-1} \rangle$. Therefore there are $\prod_{i=0}^{k-1}\left(q^k - q^i\right)$ distinct ordered k-tuples, $(\mathbf{v}_1, \ldots, \mathbf{v}_k)$. The order of $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is irrelevant, hence the number of distinct bases for V is $\frac{1}{k!}\prod_{i=0}^{k-1}\left(q^k - q^i\right)$. ¶

Corollary 54[1]. Let C be a linear code of length n over $\mathbf{F}_q$. Then, $\dim C = \log_q |C|$, in other words $|C| = q^{\dim C}$.

§

Theorem 55. Let S be a subset of $\mathbf{F}_q^n$. Then, $\dim \langle S \rangle + \dim S^{\perp} = n$.

Proof. When $\langle S \rangle = \{\mathbf{0}\}$, this is obvious. Next, consider cases where $\dim \langle S \rangle = k$, where $1 \leq k < n$. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis of $\langle S \rangle$, then we need to show that $\dim S^{\perp} = \dim \langle S \rangle^{\perp} = n - k$. Since $\mathbf{x}$ is in $S^{\perp}$ if and only if $\mathbf{v}_1 \cdot \mathbf{x} = \cdots = \mathbf{v}_k \cdot \mathbf{x} = 0$, or equivalently $A\mathbf{x} = \mathbf{0}$, where the $k \times n$ matrix A is

$$A = \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_k^T \end{bmatrix}$$

we know that the rows of A are linearly independent. Then $A\mathbf{x} = \mathbf{0}$ is a linear system of k linearly independent equations in n variables, where $n > k$, and therefore admits a solution space of dimension $n - k$. ¶

Corollary 55[1]. Let C be a linear code of length n over $\mathbf{F}_q$. Then $C^{\perp}$ is also a linear code, and $\dim C + \dim C^{\perp} = n$

Proof. This follows from Note 16 and Theorem 55 above. ¶

Theorem 56. Let C be a linear code of length n over $\mathbf{F}_q$. Then, $\left(C^{\perp}\right)^{\perp} = C$.

Proof. From Corollary 55[1], we have $\dim C + \dim C^{\perp} = n$ and $\dim C^{\perp} + \dim \left(C^{\perp}\right)^{\perp} = n$, and hence $\dim C = \dim \left(C^{\perp}\right)^{\perp}$. Let $\mathbf{c}$ be in C. Then for all $\mathbf{x}$ in C, we have $\mathbf{c} \cdot \mathbf{x} = 0$, hence $C \subseteq \left(C^{\perp}\right)^{\perp}$ and the proof. ¶

Definition 84. A *linear code* of length n over $\mathbf{F}_q$ is a subspace of $\mathbf{F}_q^n$. The *dual code* $C^\perp$ of C is the orthogonal complement of the subspace C of $\mathbf{F}_q^n$. The *dimension* of the linear code C is the dimentions of C as a vector space over $\mathbf{F}_q$, that is to say, $\dim C$. A linear code C of length n and dimension k over $\mathbf{F}_q^n$ is called a q-ary $[n,k]$-code, or an (n,q^k)-linear code. If the distance d of C is known, it is called an $[n,k,d]$-linear code. Furthermore, C is said to be *self-orthogonal* if $C \subseteq C^\perp$, and *self-dual* if $C = C^\perp$.

§

Definition 85. Let $\mathbf{x}$ be a word in $\mathbf{F}_q^n$. Then, the *Hamming weight* $w(\mathbf{x})$ of $\mathbf{x}$ is defined as the number of non-zero letters in $\mathbf{x}$. In other words, $w(\mathbf{x}) = d(\mathbf{x},\mathbf{0})$, where $\mathbf{0}$ is the zero word and $d(\mathbf{x},\mathbf{y})$ is the Hamming distance between two words $\mathbf{x}$ and $\mathbf{y}$ in $\mathbf{F}_q^n$. For each element x of $\mathbf{F}_q$, the Hamming weight may be defined as

$$w(x) = d(x,0) = \begin{cases} 1, & \text{if } x \neq 0 \\ 0, & \text{if } x = 0 \end{cases}$$

Then for $\mathbf{x} = (x_1,\ldots,x_n)$ in $\mathbf{F}_q^n$,

$$w(\mathbf{x}) = w(x_1) + \cdots + w(x_n)$$

§

Theorem 57. Let $\mathbf{x}$ and $\mathbf{y}$ be two words in $\mathbf{F}_q^n$. Then $d(\mathbf{x},\mathbf{y}) = w(\mathbf{x}-\mathbf{y})$.

Proof. For each pair of letters x and y in $\mathbf{F}_q$, we know that $d(x,y) = 0$ if and only if $x = y$, that is if and only if $x - y = 0$, or equivalently $w(x-y) = 0$. The proof follows since $w(\mathbf{x}) = \sum_{i=1}^n w(x_i)$ and $d(\mathbf{x},\mathbf{y}) = \sum_{i=1}^n d(x_i,y_i)$. ¶

Corollary 57[1]. Let q be an even positive integer. Then, for any two words $\mathbf{x}$ and $\mathbf{y}$ in $\mathbf{F}_q^n$ we have $d(\mathbf{x},\mathbf{y}) = w(\mathbf{x}+\mathbf{y})$.

Proof. The proof follows from the fact that $a = -a$ for all a in $\mathbf{F}_q$ when q is even. ¶

Theorem 58. Let $\mathbf{x}$ and $\mathbf{y}$ be two words in $\mathbf{F}_2^n$. Then, $w(\mathbf{x}) + w(\mathbf{y}) \geq w(\mathbf{x}+\mathbf{y})$.

Proof. For $\mathbf{x} = (x_1,\ldots,x_n)$ and $\mathbf{y} = (y_1,\ldots,y_n)$ in $\mathbf{F}_q^n$, let $\mathbf{x} * \mathbf{y} = (x_1y_1,\ldots,x_ny_n)$. Then, for $q = 2$ and $n = 1$,

x	y	$x*y$	$w(x)+w(y)-2w(x*y)$	w(x+y)
0	0	0	0	0
0	1	0	1	1
1	0	0	1	1
1	1	1	0	0

From this together with Definition 85 we know that $w(\mathbf{x}+\mathbf{y}) = w(\mathbf{x}) + w(\mathbf{y}) - 2w(\mathbf{x}*\mathbf{y})$ for $\mathbf{x}$ and $\mathbf{y}$ in $\mathbf{F}_2$, and thus the proof is implied. ¶

Problem 10. Prove for any prime power q and $\mathbf{x}, \mathbf{y}$ in $\mathbf{F}_q^n$, that

$$w(\mathbf{x}) + w(\mathbf{y}) \geq w(\mathbf{x} + \mathbf{y}) \geq w(\mathbf{x}) - w(\mathbf{y})$$

§

Definition 86. Let A be a matrix over $\mathbf{F}_q$. An *elementary row operation* performed on A is any one among the following.

a. interchange of two rows
b. multiplication of a row by a non-zero scalar
c. replacement of a row by its summation with a scalar multiple of another row

Two matrices are said to be *row equivalent* to each other if one is obtainable from another by a sequence of elementary row operations.

§

Definition 87. Any matrix is row equivalent to a matrix in *row echelon* (RE) form or *reduced row echelon* (RRE)† form formed by a sequence of elementary row operations done upon itself. The RRE form of any given matrix is unique, but its RE's may not be so.

§

Bibliography

San Ling and Chaoping Xing. *Coding theory, a first course.* Cambridge University Press, 2004

† The RRE form has all its leading zero of each row the only non-zero entry in its column, and its value is equal to 1.

Examples
Linear codes

14^{th} January, 2007

47. Let $q = 2$. Let $S = \{0001, 0010, 0100\}$ be a subset of a vector space V over $\mathbf{F}_2$. Find the linear span of S, $\langle S \rangle$.

Solution. Write the span as

$$\lambda_1(0001) + \lambda_2(0010) + \lambda_3(0100)$$

where $\lambda_i \in \mathbf{F}_2$, $i = 1, 2, 3$.

$\lambda_1\lambda_2\lambda_3$		vector
000	$\rightarrow$	0000
100	$\rightarrow$	0001
010	$\rightarrow$	0010
001	$\rightarrow$	0100
011	$\rightarrow$	$0010 + 0100 = 0110$
101	$\rightarrow$	$0001 + 0100 = 0101$
110	$\rightarrow$	$0001 + 0010 = 0011$
111	$\rightarrow$	$0001 + 0020 + 0100 = 0111$

#

A= **48**. Let $q = 3$ and $S = \{12101, 20110, 01122, 11010\}$. Find a basis for $C = \langle S \rangle$.

Solution. We have

$$A = \begin{pmatrix} 1 & 2 & 1 & 0 & 1 \\ 2 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 2 & 2 \\ 1 & 1 & 0 & 1 & 0 \end{pmatrix}$$

Reduce by elementary row operations our A into the row echelon form. Thus,

$$A = \begin{pmatrix} 1 & 2 & 1 & 0 & 1 \\ 2 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 2 & 2 \\ 1 & 1 & 0 & 1 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 1 & 0 & 1 \\ 0 & 2 & 2 & 1 & 1 \\ 0 & 1 & 1 & 2 & 2 \\ 0 & 2 & 2 & 1 & 2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 1 & 0 & 1 \\ 0 & 1 & 1 & 2 & 2 \\ 0 & 1 & 1 & 2 & 2 \\ 0 & 2 & 2 & 1 & 2 \end{pmatrix}$$

$$\rightarrow \begin{pmatrix} 1 & 2 & 1 & 0 & 1 \\ 0 & 1 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 1 & 0 & 1 \\ 0 & 1 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Then $\{12101, 01122, 00001\}$ is a basis for C.

#

49. Let $q = 2$ and $S = \{11101, 10110, 01011, 11010\}$ Find a basis for $C = \langle S \rangle$.

Solution. Form a matrix A according to Algorithm 4.2, and reduce it into row echelon form.

$$A = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Then the leading columns, which are those that contain the first one in each row, are 1, 2 and 4. Therefore $\{11101, 10110, 11010\}$ forms a basis for C.

#

50. Let C be the binary $[5, 3]$-linear code with the generator matrix

$$G = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Encode the message $\mathbf{u} = 101$ and find the information rate of C.

Solution. The message $\mathbf{u}$ is encoded into $\mathbf{v}$ as,

$$\mathbf{v} = \mathbf{u}G = (101) \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix} = (1 \quad 0 \quad 0 \quad 1 \quad 1)$$

#

The information rate of C is $\frac{3}{5}$

#

51. Find the cosets of the binary linear code

$$C = \{0000, 1011, 0101, 1110\}$$

Solution. We find cosets one after another, and see whether they are a new one.

$0000 + C \rightarrow 0000, 1011, 0101, 1110 \rightarrow$ I
$0001 + C \rightarrow 0001, 1010, 0100, 1111 \rightarrow$ II
$0010 + C \rightarrow 0010, 1001, 0111, 1100 \rightarrow$ III
$0011 + C \rightarrow$ IV
$0100 + C \rightarrow$ II
$0101 + C \rightarrow$ I
$0110 + C \rightarrow$ IV
$0111 + C \rightarrow$ III
$1000 + C \rightarrow 1000, 0011, 1101, 0110 \rightarrow$ IV
$1001 + C \rightarrow$ III
$1010 + C \rightarrow$ II
$1011 + C \rightarrow$ I
$1100 + C \rightarrow$ III
$1101 + C \rightarrow$ IV
$1110 + C \rightarrow$ I
$1111 + C \rightarrow$ II

Thus there are four cosets, and they are the I, II, III and IV as shown.
#

Cyclic codes

6^{th} January 2005

Definition 88. A subset S of $\mathbf{F}_q^n$ is *cyclic* if $(a_0, \ldots, a_{n-1}) \in S$ implies $(an-1, a_0, \ldots, a_{n-2}) \in S$. A linear code C is called a *cyclic code* if C is a cyclic set. The word

$$(u_{n-r}, \ldots, u_{n-1}, u_0, u_1, \ldots, u_{n-r-1})$$

is said to be obtained from the word $(u_0, \ldots, u_{n-1})$ in $\mathbf{F}_q^n$ by cyclically shifting r positions.

§

Definition 89. Let R be a ring. A nonempty subset I of R is called an *ideal* if both $a+b$ and $a-b$ belong to I and $r \cdot a$ is in I, for all a and b in I and r in R.

§

Note 17. $\mathbf{F}_q^n$, also denoted by $V(n,q)$, is the vector space of all vectors of length n over $\mathbf{F}_q$, which is also known as $GF(q)$. Here we suppose that n and q are relative primes of each other, that is to say, $(n,q) = 1$.

§

Cyclic codes can also be defined as Definition 90.

Definition 90. Let θ be a mapping such that

$$\theta : \mathbf{F}_q^n \mapsto \mathbf{F}[x]/\left(x^n - 1\right)$$

where $(x^n - 1)$, sometimes denoted by $\langle x^n - 1 \rangle$, denotes the ideal of the polynomial ring $F[x]$ generated by $x^n - 1$ by

$$\theta\left(a_0, \ldots, a_{n-1}\right) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + \langle x^n - 1 \rangle$$

for all a_i in $\mathbf{F}_q$, $0 \leq i \leq n-1$. In other words,

$$\theta : (a_0, \ldots, a_{n-1}) \mapsto a_0 + a_1 x + \cdot + a_{n-1}x^{n-1}$$

§

Note 18. $\mathbf{F}[x]/\left(x^n - 1\right)$ is also a vector space over $\mathbf{F}$ is an $\mathbf{F}_q$-linear transformation of vector spaces over $\mathbf{F}_q$. In fact it is a vector space isomorphism. We could identify $\mathbf{F}_q^n$ with $\mathbf{F}_q[x]/\left(x^n - 1\right)$, and a vector $\mathbf{u} = (a_0, \ldots, a_{n-1})$ with the polynomial $a(x) = \sum_{i=0}^{n-1} a_i x^i$

§

Theorem 59. Let θ be the mapping defined in Definition 90, and let C be a linear code of length n over $\mathbf{F}_q$. Then C is a cyclic code if and only if $\theta(C)$ is an ideal in the quotient ring $\mathbf{F}[x]/\left(x^n-1\right)$.

Proof. Since C is a linear code of length n over $\mathbf{F}_q$, it is a subspace of the vector space $\mathbf{F}_q^n$, and $\theta(C)$ a subspace of $\mathbf{F}[x]/\left(x^n-1\right)$. Let $(a_0,\ldots,a_{n-1})$ be in C. Then $(a_{n-1},a_0,\ldots,a_{n-2})$ is in C if and only if

$$a_{n-1}+a_0x+\cdots+a_{n-2}x^{n-1}+\langle x^n-1\rangle =$$

$$x\left(a_0+a_1x+\cdots+a_{n-1}x^{n-1}\right)+\langle x^n-1\rangle$$

is in $\theta(C)$. ¶

Example 32. Recall that a $[n,k,d]$-linear code is a code of distance d, length n and the number of elements in the bases k. Then the binary $[3,2,2]$-linear code $\{000,110,101,011\}$ is a cyclic code and $\theta(C)=\left\{0,1+x,1+x^2,x+x^2\right\}$ is a subset of $\mathbf{F}_2[x]/\left(x^3-1\right)$. Moreover, $\theta(C)$ is an ideal.

Example 33. The set of all the integers divisible by a fixed positive integer m is an ideal of $\mathbf{Z}$. All the polynomials in the polynomial ring $\mathbf{F}_q[x]$ that are divisible by a fixed, non-zero polynomial $f(x)$ form an ideal.

Definition 91. An ideal I of a ring R is called a *principal ideal* if there exists an element g in I such that

$$I=\langle g\rangle=\{gr:r\in R\}$$

The element g is called a *generator* of I, and I is said to be generated by g. A ring is called a *principal ideal ring* if all its ideals are principal.

§

Bibliography

Raymond Hill. *A first course in coding theory.* Clarendon, 1986

San Ling and Chaoping Xing. *Coding theory, a first course.* Cambridge University Press, 2004

L R Vermani. *Elements of algebraic codng theory.* Chapman & Hall, 1996

Examples
Cyclic codes
14^{th} January, 2007

52. Show whether the sets $\{(0,1,1,2),(2,0,1,1),(1,2,0,1),(1,1,2,0)\} \subset \mathbf{F}_3^4$ and $\{11111\} \subset \mathbf{F}_2^5$ are cyclic codes.

Solution. Both are cyclic sets since $(a_{n-1}, a_0, a_1, \ldots, a_{n-2})$ is in S for all $(a_0, \ldots, a_{n-1})$ in S. But for the former set $0112 + 2011 = 2120$, which is not in the set, therefore the set is not a cyclic code.

#

Similarly for the latter, since $11111 + 11111 = 00000$, not in the set. Hence this set is also not a cyclic code.

#

53. Show that in a the ring $\mathbf{F}_2[x]/\left(x^3-1\right)$ the subset

$$I = \left\{0, 1+x, x+x^2, 1+x^2\right\}$$

is an ideal.

Solution. A non-empty subset I of a ring R is an ideal if both $a+b$ and $a-b$ belong to I for all a and b in I, and if $r \cdot a \in I$ for all r in R and a in I. Here $q = 2$, from which we know that $a + b = a - b$. Since

$$\left.\begin{array}{cc} 0+1+x & =1+x \\ 0+x+x^2 & =x+x^2 \\ 1+x+x+x^2 & =1+x^2 \\ 1+x+1+x^2 & =x+x^2 \\ x+x^2+x+x^2 & =0 \\ x+x^2+1+x^2 & =1+x \end{array}\right\} \in I$$

Next, the ring R being

$$\mathbf{F}_2[x]/\left(x^3-1\right) = \left\{0, 1, x, 1+x, x^2, 1+x^2, x+x^2, 1+x+x^2\right\}$$

and its subset to consider

$$I = \left\{0, 1+x, x+x^2, 1+x^2\right\}$$

the multiplication table of $r \cdot a$ for all $r \in R$ and $a \in I$ is

$\cdot$	0	$1+x$	$1+x^2$	$x+x^2$
0	0	0	0	0
1	0	$1+x$	$1+x^2$	$x+x^2$
x	0	$x+x^2$	$1+x$	$1+x^2$
$1+x$	0	$1+x^2$	$x+x^2$	$1+x$
x^2	0	$1+x^2$	$x+x^2$	$1+x$
$1+x^2$	0	$x+x^2$	$1+x$	$1+x^2$
$x+x^2$	0	$1+x$	$1+x^2$	$x+x^2$
$1+x+x^2$	0	0	0	0

Table 5. $r \cdot a$ for all $r \in R$ and $a \in I$

From Table 5 we can see that $r \cdot a$ is in I. Hence we conclude that the subset I of R is an ideal.

#

54. Find how many binary cyclic codes of length 6 there are. Give one cyclic code as an example.

Solution. First we factorise the polynomial $x^6 - 1 \in \mathbf{F}_2[x]$ thus,

$$x^6 - 1 = (1+x)^2 \left(1 + x + x^2\right)^2$$

Next, list all the monic divisors of $x^6 - 1$,

$$1, \quad 1+x, \quad 1+x+x^2, \quad (1+x)^2, \quad (1+x)\left(1+x+x^2\right),$$

$$(1+x)^2\left(1+x+x^2\right), \quad \left(1+x+x^2\right)^2, \quad (1+x)\left(1+x+x^2\right)^2, \quad \left(1+x^6\right)$$

Since the nuber of these is nine, there are nine binary codes of length 6. We can then write down all these cyclic codes based on the map π, for example the one corresponding to the polynomial $\left(1+x+x^2\right)^2$ is found by

$$\left(1+x+x^2\right)^2 = \left(1+x+x^2\right)\left(1+x+x^2\right) = 1 + x^2 + x^4$$

$$\begin{array}{lll} 0 \cdot \left(1+x^2+x^4\right) & = 0 & \to \quad 000000 \\ 1 \cdot \left(1+x^2+x^4\right) & = 1+x^2+x^4 & \to \quad 101010 \\ x \cdot \left(1+x^2+x^4\right) & = x+x^3+x^5 & \to \quad 010101 \end{array}$$

Then all the additions among these give another word, $010101 + 101010 = 111111$. Hence the cyclic code is

$$\{000000, 101010, 010101, 111111\}$$

#

55. Based on the factorisation $x^7 - 1 = (1+x)\left(1+x^2+x^3\right)\left(1+x+x^3\right) \in \mathbf{F}_2[x]$. Find the number of different binary $[7,3]$-cyclic codes,

Solution. First we list all the 8 monic divisors of $x^7 - 1$,

$$1, \quad 1+x, \quad \left(1+x^2+x^3\right), \quad \left(1+x+x^3\right),$$

$$(1+x)\left(1+x^2+x^3\right), \quad (1+x)\left(1+x+x^3\right),$$

$$\left(1+x^2+x^3\right)\left(1+x+x^3\right), \quad (1+x)\left(1+x^2+x^3\right)\left(1+x+x^3\right)$$

We need a code for which the dimention is $k = 3$. Write all the bases of each of these monic divisors.

$1, x, x^2, \ldots, x^6$ $\rightarrow$ k=7
$1 + x, x + x^2, x^2 + x^3, \ldots, x^5 + x^6$ $\rightarrow$ $k = 6$
$\left(1 + x^2 + x^3\right)$ $\rightarrow$ $k = 4$
$\left(1 + x + x^3\right)$ $\rightarrow$ $k = 4$
$(1 + x)\left(1 + x^2 + x^3\right), x^i(1 + x)\left(1 + x^2 + x^3\right)\ (i = 1, 2)$ $\rightarrow$ $k = 3$
$(1 + x)\left(1 + x + x^3\right), x^i(1 + x)\left(1 + x + x^3\right)\ (i = 1, 2)$ $\rightarrow$ $k = 3$
$\left(1 + x^2 + x^3\right)\left(1 + x + x^3\right)$ $\rightarrow$ $k = 1$
$(1 + x)\left(1 + x^2 + x^3\right)\left(1 + x + x^3\right)$ $\rightarrow$ $k = 0$

From this, we see that there are two such divisors which have $k = 3$. This means that there are exactly two different binary $[7, 3]$-cyclic codes.

For the first one, $(1 + x)\left(1 + x^2 + x^3\right) = 1 + x + x^2 + x^4$.

$$\begin{aligned} 0 \cdot \left(1 + x + x^2 + x^4\right) &\rightarrow 0000000 \\ 1 \cdot \left(1 + x + x^2 + x^4\right) &\rightarrow 1110100 \\ x \cdot \left(1 + x + x^2 + x^4\right) &\rightarrow 0111010 \\ x^2 \cdot \left(1 + x + x^2 + x^4\right) &\rightarrow 0011101 \end{aligned}$$

And the pairwisely exhaustive additions yield,

$$\begin{aligned} 1110100 + 0111010 &= 1001110 \\ 1110100 + 0011101 &= 1101001 \\ 0111010 + 0011101 &= 0100111 \\ 1110100 + 0111010 + 0011101 &= 1010011 \end{aligned}$$

Therefore,

$$\left\langle (1 + x)\left(1 + x^2 + x^3\right)\right\rangle =$$

$$\{0000000, 1110100, 0111010, 0011101, 1001110, 0100111, 1010011, 1101001\}$$

#

Similarly for the second one,

$$\left\langle (1 + x)\left(1 + x + x^3\right)\right\rangle =$$

$$\{0000000, 1011100, 0101110, 0010111, 1001011, 1100101, 1110010, 0111001\}$$

#

Exercises on cyclic codes

14^{th} January, 2007

56. *Determine* all the binary cyclic codes of length 9.
57. *Is* $x^6 + x^3 + 1$ irreducible over $\mathbf{F}_2$? If it is, then *generate* using the same the binary code of length 7 and dimension 3.

Reference

Raymond Hill. *A first course in coding theory.* Clarendon, 1986

San Ling and Chaoping Xing. *Coding theory, a first course.* Cambridge University Press, 2004

L R Vermani. *Elements of algebraic codng theory.* Chapman & Hall, 1996

Goppa codes

13^{th} January 2006

Definition 92. A linear code with parameter $[n, k, d]$ such that $k+d = n+1$ is called a *maximum distance separable* (MDS) code.

§

Theorem 60. Let C be a linear code over $\mathbf{F}_q$ with parameters $[n, k, d]$. Let G be a generator matrix, and H a parity matrix, for C. Then, the following statements are equivalent.

a. C is an MDS code,
b. every set of $n - k$ columns of H is linearly independent,
c. every set of k columns of G is linearly independent,
d. $C^{\perp}$ is an MDS code.

§

Definition 93. An MDS code C over $\mathbf{F}_q$ is said to be *trivial* if and only if C satisfies one of the following cases.

a. $C = \mathbf{F}_q^n$,
b. C is equivalent to the code generated by $\mathbf{1} = (1, \ldots, 1)$,
c. C is equivalent to the dual of the code generated by $\mathbf{1}$. C is said to be *nontrivial* if it is not trivial.

§

The class of Bose, Chaudhuri and Hocquenghem (BCH) codes is a generalisation of the Hamming codes for multiple-error correction. Binary BCH codes were introduced by A Hocquenghem (1959) and then independently by R C Bose and D K Ray-Chaudhuri (1960). D Gorenstein and N Zierler (1961) generalised the binary BCH codes to q-ary ones. The class of Reed-Solomon (RS) codes is a subclass of BCH codes introduced by I S Reed and G Solomon (1960). Goppa codes, a generalisation of BCH codes introduced by V D Goppa (1970 and 1971), are used also in cryptography some examples of which are the McEliece- and the Niederreiter cryptosystems. The Goppa codes are in turn a subclass of alternant codes, which was introduced by H J Helgert in 1974.

Theorem 61. Let $(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$ be an arbitrary ordering of the $n = 2^m - 1$ non-zero elements of $\mathbf{F}_{2^m}$. Than a word $\mathbf{c} = \{c_0, \ldots, c_{n-1}\}$ is a code word of BCH code if and only if $\sum_{i=0}^{n-1} c_i \alpha_i^j = 0$, where $j = 1, 2, \ldots, 2t$.

§

Definition 94. A q-ary Reed-Solomon (RS) code is a q-ary BCH code of length $q - 1$ generated by

$$g(x) = \left(x - \alpha^{a+1}\right)\left(x - \alpha^{a+2}\right)\cdots\left(x - \alpha^{a+\delta-1}\right)$$

where α is a primitive element of $\mathbf{F}_q$, $a \geq 0$ and $2 \leq \delta \leq q - 1$.

§

Theorem 62. Reed-Solomon codes are MDS. This means that a q-ary Reed-Solomon code of length $q-1$ generated by $g(x) = \prod_{i=a+1}^{a+\delta-1} \left(x - \alpha^i\right)$ is a $\{q-1, q-\delta, \delta\}$-cyclic code for any $2 \leq \delta \leq q-1$.

§

Theorem 63. Let C be a q-ary RS code generated by $g(x) = \prod_{i=1}^{\delta-1} \left(x - \alpha^i\right)$, where $2 \leq \delta \leq q-1$. Then the extended code $\overline{C}$ is also MDS.

§

Theorem 64. Let α be a primitive element of the finite field $\mathbf{F}_q$. Let $q-1 \geq \delta \geq 2$. The narrow-sense q-ary RS code with generator polynomial

$$g(x) = (x-\alpha)\left(x-\alpha^2\right)\cdots\left(x-\alpha^{\delta-1}\right)$$

is equal to

$$\left\{\left(f(1), f(\alpha), f(\left(\alpha^2\right)), \ldots, f\left(\alpha^{q-2}\right)\right) : f(x) \in \mathbf{F}_q[x] \text{ and } \deg\left(f(x)\right) < q-\delta\right\}$$

§

Theorem 65. Let α be a primitive element of $\mathbf{F}_q$, and let $q-1 \geq \delta \geq 2$. The matrix

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \alpha & \alpha^2 & \cdots & \alpha^{q-2} \\ 1 & \alpha^2 & \alpha^4 & \cdots & \alpha^{2(q-2)} \\ \vdots & & \vdots & \ddots & \vdots \\ 1 & \alpha^{q-\delta-1} & \alpha^{2(q-\delta-1)} & \cdots & \alpha^{(q-\delta-1)(q-2)} \end{pmatrix}$$

is a generator matrix for the RS code generated by the polynomial

$$g(x) = (x-\alpha)\left(x-\alpha^2\right)\cdots\left(x-\alpha^{\delta-1}\right)$$

§

Definition 95. Let $n \leq q$. Let $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, where α_i, $1 \leq i \leq n$, are distinct elements of $\mathbf{F}_q$. Let $\mathbf{v} = (v_1, \ldots, v_n)$, where $v_i \in \mathbf{F}_q^*$ for all $1 \leq i \leq n$. The *generalised Reed-Solomon* code $GRS_k(\alpha, v)$ is defined as

$$\{v_1 f(\alpha_1), v_2 f(\alpha_2), \ldots, v_n f(\alpha_n) : f(x) \in \mathbf{F}_q[x] \quad \text{and} \quad \deg\left(f(x)\right) < k \leq n\}$$

.

§

Theorem 66. The dual of the generalised Reed-Solomon code $GRS_k(\alpha, vb)$ over $\mathbf{F}_q$ of length n is $GRS_{n-k}\left(\alpha, \mathbf{v}^9 0\right)$ for some $\mathbf{v}^9 0 \in \left(\mathbf{F}_q^*\right)^n$.

§

Theorem 67.

$$\begin{pmatrix} v_1^9 0 & v_2^9 0 & \cdots & v_n^9 0 \\ v_1^9 0\alpha_1 & v_2^9 0\alpha_2 & \cdots & v_n^9 0\alpha_n \\ v_1^9 0\alpha_1^2 & v_2^9 0\alpha_2^2 & \cdots & v_n^9 0\alpha_n^2 \\ \vdots & & \ddots & \vdots \\ v_1^9 0\alpha_1^{n-k-1} & v_2^9 0\alpha_2^{n-k-1} & \cdots & v_n^9 0\alpha_n^{n-k-1} \end{pmatrix}$$

§

Definition 96. An *alternant* code $A_k\left(\alpha, \mathbf{v}^9 0\right)$ over the finite field $\mathbf{F}_q$ is the subfield subcode $GRS_k(\alpha, \mathbf{v})|_{\mathbf{F}_q}$, where $GRS_k(\alpha, \mathbf{v})$ is a generalised RS code over $\mathbf{F}_{q^m}$, for some $m \geq 1$.

§

Theorem 68. The alternant code $A_k\left(\alpha, \mathbf{v}^9 0\right)$ has parameters $\left[n, k^9 0, d\right]$, where $mk - (m-1)n \leq k^9 0 \leq k$ and $d \geq n - k + 1$.

§

Theorem 69. The dual of the alternant code $A_k\left(\alpha, \mathbf{v}^9 0\right)$ is

$$\underset{\mathbf{F}_{q^m}/\mathbf{F}_q}{\mathrm{Tr}}\left(GRS_{n-k}\left(\alpha, \mathbf{v}^9 0\right)\right)$$

§

Theorem 70. Given any positive integers n, h, δ and m. If

$$\sum_{w=0}^{\delta-1}(q-1)^w \binom{n}{w} < (q^m - 1)^{\left\lfloor \frac{n-h}{m} \right\rfloor}$$

then there exists an alternant code $A_k\left(\alpha, \mathbf{v}^9 0\right)$ over $\mathbf{F}_q$, which is the subfield subcode of a generalised RS code over $\mathbf{F}_{q^m}$, having parameters $\left\{n, k^9 0, d\right\}$, where $k^9 0 \geq h$ and $d \geq \delta$.

§

Definition 97. Let $g(z)$ be a polynomial in $\mathbf{F}_{q^m}[z]$. Let $L = \{\alpha_1, \ldots, \alpha_n\}$ be a subset of $\mathbf{F}_{q^m}$ such that $L \cap \{\text{zeros of } g(z)\} = \emptyset$. Let $R_c(z) = \sum_{i=1}^n \frac{c_i}{z - \alpha_i}$ for $\mathbf{c} = (c_1, \ldots, c_n) \in \mathbf{F}_q^n$. Then, the *Goppa code* $\Gamma(L, g)$ is defined as

$$\Gamma(L, g) = \left\{\mathbf{c} \in \mathbf{F}_q^n : R_c(z) \cong 0 \left(\mathrm{mod}\, g(z)\right)\right\}$$

The polynomial $g(z)$ is called the *Goppa polynomial.* The Goppa code $\Gamma(L, g)$ is said to be *irreducible* if $g(z)$ is irreducible.

§

Theorem 71. A word is a code word of the Goppa code, that is to say, $\mathbf{c} \in \Gamma(L, g)$ if and only if

$$\sum_{i=1}^{n} \frac{g(z) - g\left(\alpha_i\right)}{z - \alpha_i} g\left(\alpha_i\right)^{-1} = 0$$

§

Theorem 72. Given a Goppa polynomial $g(z)$ of degree t and

$$L = \{\alpha_1, \ldots, \alpha_n\}$$

we have $\Gamma(L, g) = \left\{\mathbf{c} \in \mathbf{F}_q^n : \mathbf{c}H^T = \mathbf{0}\right\}$, where

$$H = \begin{pmatrix} g(\alpha_1)^{-1} & \cdots & g(\alpha_n)^{-1} \\ \alpha_1 g(\alpha_1)^{-1} & \cdots & \alpha_n g(\alpha_n)^1 \\ \vdots & \ddots & \vdots \\ \alpha_1^{t-1} g(\alpha_1)^{-1} & \cdots & \alpha_n^{t-1} g(\alpha_n)^{-1} \end{pmatrix}$$

§

Theorem 73. Given a Goppa polynomial $g(z)$ of degree t and

$$L = \{\alpha_1, \ldots, \alpha_n\}$$

the Goppa code $\Gamma(L, g)$ is the alternant code $A_{n-1}\left(\alpha, \mathbf{v}^9 0\right)$, where $\alpha = (\alpha_1, \ldots, \alpha_n)$ and

$$\mathbf{v}^9 0 = \left(g(\alpha_1)^{-1}, \ldots, g(\alpha_n)^{-1}\right)$$

§

Theorem 74. The Goppa code $\Gamma(L, g)$ is $GRS_{n-t}(\alpha, \mathbf{v})|_{\mathbf{F}_q}$, where $\mathbf{v} = (v_1, \ldots, v_n)$ and

$$\frac{v_i = g(\alpha_i)}{\prod_{j \neq i}((\alpha_i - \alpha_j))}$$

for all $1 \leq i \leq n$.

§

Theorem 75. Given a Goppa polynomial $g(z)$ of degree t and

$$L = \{\alpha_1, \ldots, \alpha_n\}$$

the Goppa code $\Gamma(L, g)$ is a linear code over $\mathbf{F}_q$ with parameters $[n, k, d]$, where $k \geq n - mt$ and $d \geq t + 1$.

§

Theorem 76. The dual of the Goppa code $\Gamma(L, g)$ is the trace code $\mathrm{Tr}_{\mathbf{F}_{q^m}/\mathbf{F}_q}\left(GRS_t\left(\alpha, \mathbf{v}^9 0\right)\right)$, where $\mathbf{v}^9 0 = \left(g(\alpha_1)^{-1}, \ldots, g(\alpha_n)^{-1}\right)$.

§

Theorem 77. Let $q = 2$. Given a polynomial $g(z)$, let $\tilde{g}(z)$ represent the lowest degree perfect square polynomial that is divisible by $g(z)$, and let $\tilde{t}$

the degree of $\tilde{g}(z)$. For a vector $\mathbf{c} = (c_1, \ldots, c_n) \in \mathbf{F}_q^n$ of weight w, where $c_{i_1} = \cdots = c_{i_w} = 1$, let

$$f_c(z) = \prod_{j=1}^{w} (z - \alpha_{i_j})$$

The derivative of $f_c(z)$ is

$$f_c^9 0(z) = \sum_{l=1}^{w} \prod_{j \neq l} (z - \alpha_{i_j})$$

Then, $\mathbf{c} \in \mathbf{F}_2^n$ belongs to $\Gamma(L, g)$ if and only if $\tilde{g}(z)$ divides $f_c^9 0(z)$. Consequently, the minimum distance d of $\Gamma(L, g)$ satisfies $d \geq \tilde{t} + 1$. If $g(z)$ has no multiple root, that is $g(z)$ is a separable polynomial, then $d \geq 2t + 1$.

§

Theorem 78. There exists a q-ary Goppa code $\Gamma(L, g)$, where $g(z)$ is an irreducible polynomial in $\mathbf{F}_{q^m}[z]$ of degree t and $L = \mathbf{F}_{q^m}$ of parameters $[q^m, k, d]$ such that $k \geq q^m - mt$, provided that

$$\sum_{w=t+1}^{d-1} \left\lfloor \frac{w-1}{t} \right\rfloor (q-1)^w \binom{q^m}{w} < \frac{1}{t} q^{mt} \left(1 - (t-1)q^{-\frac{mt}{2}}\right)$$

§

Bibliography

R C Bose and D K Ray-Chaudhuri. On a class of error-correcting binary group codes. *Inform. Control.* **3**, 68–79, 1960

Raymond Hill. *A first course in coding theory.* Clarendon, 1986

V D Goppa. A new class of linear error-correcting codes. *Probl. Peredach. Inform.* **6**, 3, 24–30, 1970

V D Goppa. Rational representation of codes and (L, g) codes. *Probl. Peredach. Inform.* **7**, 3, 41–9, 1971

D Gorenstein and N Zierler. A class of cyclic linear error-correcting codes in p^m symbols. *J. Soc. Ind. App. Math.* **9**, 107–214, 1961

H J Helgert. Alternant codes. *Information and Control.* **26**, 369–80, 1974

A Hocquenghem. Codes correcteurs d'erreurs. *Chiffres.* **2**, 147–56, 1959

San Ling and Chaoping Xing. *Coding theory, a first course.* Cambridge University Press, 2004

I S Reed and G Solomon. Polynomial codes over certain finite fields. *J.Soc.Ind. App. Math.* **8**, 300–4, 1960

Exercises for Goppa codes

13^{th} January 2006

58. Let

$$G = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \\ v_1\alpha_1 & v_2\alpha_2 & \cdots & v_n\alpha_n \\ \vdots & \vdots & \ddots & \vdots \\ v_1\alpha_1^{k-1} & v_2\alpha_2^{k-1} & \cdots & v_n\alpha_n^{k-1} \end{pmatrix}$$

be a generator matrix for the generalised RS code $GRS_k(\alpha, \mathbf{v})$. Let C be the code with generator matrix $\left(G|\mathbf{u}^T\right)$, where $\mathbf{u} = (0, \ldots, 0, u)$, for some $u \in \mathbf{F}_q^*$. Let $\mathbf{v}^9 0 = \left(v_1^9 0, \ldots, v_n^9 0\right)$ be such that $GRS_{n-k}\left(\alpha, \mathbf{v}^9 0\right)$ is the dual of $GRS_k(\alpha, \mathbf{v})$.

i. Show that there is some $w \in \mathbf{F}_q^*$ such that

$$\sum_{i=1}^{n} v_i v_i^9 0 \alpha_i^{n-1} + uw = 0$$

ii. Show that

$$H^9 0 = \begin{pmatrix} v_1^9 0 & v_2^9 0 & \cdots & v_n^9 0 & 0 \\ v_1^9 0\alpha_1 & v_2^9 0\alpha_2 & \cdots & v_n^9 0\alpha_n & 0 \\ v_1^9 0\alpha_1^2 & v_2^9 0\alpha_2^2 & \cdots & v_n^9 0\alpha_n^2 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ v_1^9 0\alpha_1^{n-k} & v_2^9 0\alpha_2^{n-k} & \cdots & v_n^9 0\alpha_n^{n-k} & w \end{pmatrix}$$

is a parity-check matrix for C.

iii. Prove that C is an MDS code.

59. Let n be odd and let $\mathbf{F}_{2^m}$ be an extension of $\mathbf{F}_2$ containing all the n^{th} roots of 1. Let α be a primitive n^{th} root of 1 in $\mathbf{F}_{2^m}$ and let $L = \left\{1, \alpha, \ldots, \alpha^{n-1}\right\}$. For $\mathbf{c} = (c_0, \ldots, c_{n-1}) \in \mathbf{F}_2^n$, let

$$R_c(z) = \sum_{i=0}^{n-1} c_i x^i$$

and let $\hat{c}(z)$ be its Mattson-Solomon polynomial.

i. Show that $\hat{c}(z) = \left(z\left(z^n + 1\right) R_c(z) \left(\bmod z^n - 1\right)\right)$ and

$$R_c(z) = \sum_{i=0}^{n-1} \frac{\hat{c}\left(\alpha^i\right)}{z + \alpha^i}$$

ii. Show that the Goppa code $\Gamma(L, g)$ is equal to

$$\Gamma(L, g) = \left\{\mathbf{c} \in \mathbf{F}_2^n : \left(z^{n-1}\hat{c}(z) \left(\bmod z^n - 1\right)\right) \cong 0 (\bmod\, g(z))\right\}$$

Hint: For (i), show that $z\left(z^n+1\right)R_c(z) = \sum_{i=0}^{n-1} c_i z \prod_{j\neq i}\left(z+\alpha^j\right)$. Then show that $\left(z\prod_{j\neq i}\left(z+\alpha^j\right)\left(\mathrm{mod}\, z^n-1\right)\right) = \sum_{j=0}^{n-1}\alpha^{-ij}z^j$ by multiplying both sides by $z+\alpha^i$. For (ii), show that $\mathbf{c}\in\Gamma(L,g)$ if and only if $\sum_{i=0}^{n-1} c_i \prod_{j\neq i}\left(z+\alpha^j\right) \cong 0(\mathrm{mod}\, g(z))$, and then use (i).)

Reference

San Ling and Chaoping Xing. *Coding theory, a first course.* Cambridge University Press, 2004

MDS code

20^{th} January 2006

Theorem 79. Given a redundancy r and a minimum distance d. An $[n, n-r, d]$-code satisfies $d \leq r+1$.

§

Definition 98. A linear $[n, k, d]$ code over F with $d = n-k+1$ is called a *maximum distance separable* (MDS) code.

In other words, an MDS is a $[n, n-r, r+1]$-code.

§

Theorem 80. Suppose $2 \leq r \leq q$. Let $a_1, \ldots, a_{q-1}$ be the non-zero elements of $GF(q)$. Then the matrix

$$H = \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 & 0 & \cdots & 0 \\ a_1 & a_2 & \cdots & a_{q-1} & 0 & 1 & \cdots & 0 \\ a_1^2 & a_2^2 & \cdots & a_{q-1}^2 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ a_1^{r-1} & a_2^{r-1} & \cdots & a_{q-1}^{r-1} & 0 & & \cdots & 1 \end{bmatrix}$$

is the parity check matrix of an MDS $q+1, q+1-r, r+1$ code. Equivalently, the columns of H form a $(q+1)$-arc in $PG(r-1, q)$.

§

Theorem 81. Let C be a linear $[n, k, d]$ code over a field F of q elements, where q is a prime power with a parity check matrix H. Then C has a code word of weight $w \leq l$ if and only if l columns of H are linearly dependent.

§

Theorem 82. Let C be a linear $[n, k, d]$ code over F with a parity check matrix H. Then C is an MDS code if and only if every $n-k$ columns of H are linearly independent.

§

Theorem 83. If a linear $[n, k, d]$ code C is MDS, then so is its dual $C^{\perp}$.

§

Corollary 83[1]. Let C be an $[n, k, d]$ linear code over $F = GF(q)$. Then the following statements are equivalent.

a. C is MDS
b. Every k columns of a generator matrix G of C are linearly independent
c. Every $n-k$ columns of a parity check matrix H of C are linearly independent

§

Problem 11. Show that linear $[n, 1, n]$, $[n, n-1, 2]$ and $[n, n, 1]$ codes exist over any finite field F.

§

Definition 99. We call *trivial MDS codes* the $[n, 1, n]$, $[n, n-1, 2]$ and $[n, n, 1]$ codes.

§

Theorem 84. The only binary MDS codes are the trivial ones.

§

Definition 100. A square matrix is said to be *non-singular* if its columns are linearly independent. Given any matrix A, a $s \times s$ *square submatrix* of A is a $s \times s$ matrix consisting of the entries from some s rows and s column of A.

§

Theorem 85. Let C be an $[n, k, -]$ code with parity check matrix $H = (A \quad I_{n-k})$. Then C is an MDS code if and only if every square submatrix of A is non-singular.

Proof. Let B_r be a square submatrix of A which rests upon the i_1^{th}, i_2^{th}, ..., i_r^{th} rows of A with $i_1 < i_2 < \cdots < i_r \leq n-k$. Let M_r be the square submatrix of H of order $n-k$ having the columns of A parts which occur in B_r and the ramaining $n-k-r$ columns from I_{n-k} that are not the i_1^{th}, i_2^{th}, ..., i_r^{th} columns. Thus we could always find the determinant by pivotting on the ones in the columns I_j, $j \neq i_1, i_2, \cdot, i_r$ successively. Then $\det M_r = \mathrm{p}\det B_r$. Therefore B_r is non-singular if and only if M_r is. Hence every $n-k$ columns of H are linearly independent if and only if every square submatrix of A is non-singular. ¶

Theorem 86. Let C be an $[n, k, -]$ code with generator matrix $G = (I_k \quad A)$. Then C is an MDS code if and only if every square submatrix of A is non-singular.

§

Theorem 87. Let C be an $[n, k, d]$ MDS code. Then any k symbols of the code words may be taken as message symbols.

§

Theorem 88. Let C be an $[n, k, d]$ code over $GF(q)$. Then C is an MDS code if and only if C has a minimum distance code word with non-zero entries in any d coordinates.

§

Corollary 88[1]. The number of code words of weight $n-k+1$ in an $[n, k, d]$ MDS code over $GF(q)$ is

$$(q-1)\binom{n}{n-k+1}$$

§

Problem 12. Given k and q, find the largest value, $m(k,q)$, of n such that $[n,k,n-k+1]$ MDS code exists over $GF(q)$.

§

Because of Theorem 83, Problem 12 is equivalent to Problem 13.

Problem 13. Given k and q, find the largest n for which there is a $k \times n$ matrix over $GF(q)$, every k columns of which are linearly independent.

§

Problem 14. Given a k-dimensional vector space V over $GF(q)$, what is the order of a largest subset of V every k vectors of which form a basis of the same?.

§

Theorem 89. For any prime power q, we have $m(2,q) = q+1$.

§

Theorem 90.

$$m(k,q) = k+1$$

for $q \leq k$.

§

Bibliography

Raymond Hill. *A first course in coding theory.* Clarendon, 1986
L R Vermani. *Elements of algebraic codng theory.* Chapman & Hall, 1996

Examples
MDS codes

14^{th} January, 2007

60. Consider the matrix

$$A = \begin{bmatrix} 3 & 5 & 6 & 2 & 1 \\ 4 & 4 & 6 & 1 & 3 \\ 2 & 5 & 2 & 1 & 6 \end{bmatrix}$$

over $GF(7)$. Show whether minimum distance separable (MDS) codes can be obtained from A. If they could, find two such codes and give either a generator matrix or a parity check matrix for each of them. Then give the code words and encoding functions for each.

Solution. Examine the values of determinant of all submatrices of A. We have, $\begin{vmatrix} 3 & 5 \\ 4 & 4 \end{vmatrix} = 6$, $\begin{vmatrix} 3 & 6 \\ 4 & 6 \end{vmatrix} = 1$, $\begin{vmatrix} 3 & 2 \\ 4 & 1 \end{vmatrix} = 2$, $\begin{vmatrix} 3 & 1 \\ 4 & 3 \end{vmatrix} = 5$, $\begin{vmatrix} 5 & 6 \\ 4 & 6 \end{vmatrix} = 6$, $\begin{vmatrix} 5 & 2 \\ 4 & 1 \end{vmatrix} = 4$, $\begin{vmatrix} 5 & 1 \\ 4 & 3 \end{vmatrix} = 4$, $\begin{vmatrix} 6 & 2 \\ 6 & 1 \end{vmatrix} = 1$, $\begin{vmatrix} 6 & 1 \\ 6 & 3 \end{vmatrix} = 5$, $\begin{vmatrix} 2 & 1 \\ 1 & 3 \end{vmatrix} = 5$, $\begin{vmatrix} 3 & 5 \\ 2 & 5 \end{vmatrix} = 5$, $\begin{vmatrix} 3 & 6 \\ 2 & 2 \end{vmatrix} = 1$, $\begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix} = 6$, $\begin{vmatrix} 3 & 1 \\ 2 & 6 \end{vmatrix} = 2$, $\begin{vmatrix} 5 & 6 \\ 5 & 2 \end{vmatrix} = 1$, $\begin{vmatrix} 5 & 2 \\ 5 & 1 \end{vmatrix} = 2$, $\begin{vmatrix} 5 & 1 \\ 5 & 6 \end{vmatrix} = 4$, $\begin{vmatrix} 6 & 2 \\ 2 & 1 \end{vmatrix} = 2$, $\begin{vmatrix} 6 & 1 \\ 2 & 6 \end{vmatrix} = 6$, $\begin{vmatrix} 2 & 1 \\ 1 & 6 \end{vmatrix} = 4$, $\begin{vmatrix} 4 & 4 \\ 2 & 5 \end{vmatrix} = 5$, $\begin{vmatrix} 4 & 6 \\ 2 & 2 \end{vmatrix} = 3$, $\begin{vmatrix} 4 & 1 \\ 2 & 1 \end{vmatrix} = 2$, $\begin{vmatrix} 4 & 3 \\ 2 & 6 \end{vmatrix} = 4$, $\begin{vmatrix} 4 & 6 \\ 5 & 2 \end{vmatrix} = 6$, $\begin{vmatrix} 4 & 1 \\ 5 & 1 \end{vmatrix} = 6$, $\begin{vmatrix} 4 & 3 \\ 5 & 6 \end{vmatrix} = 2$, $\begin{vmatrix} 6 & 1 \\ 2 & 1 \end{vmatrix} = 4$, $\begin{vmatrix} 6 & 3 \\ 2 & 6 \end{vmatrix} = 2$, $\begin{vmatrix} 1 & 3 \\ 1 & 6 \end{vmatrix} = 3$, $\begin{vmatrix} 3 & 5 & 6 \\ 4 & 4 & 6 \\ 2 & 5 & 2 \end{vmatrix} = 5$, $\begin{vmatrix} 3 & 5 & 2 \\ 4 & 4 & 1 \\ 2 & 5 & 1 \end{vmatrix} = 4$, $\begin{vmatrix} 3 & 5 & 1 \\ 4 & 4 & 3 \\ 2 & 5 & 6 \end{vmatrix} = 5$, $\begin{vmatrix} 3 & 6 & 2 \\ 4 & 6 & 1 \\ 2 & 2 & 1 \end{vmatrix} = 6$, $\begin{vmatrix} 3 & 6 & 1 \\ 4 & 6 & 3 \\ 2 & 2 & 6 \end{vmatrix} = 6$, $\begin{vmatrix} 3 & 2 & 1 \\ 4 & 1 & 3 \\ 2 & 1 & 6 \end{vmatrix} = 3$, $\begin{vmatrix} 5 & 6 & 2 \\ 4 & 6 & 1 \\ 5 & 2 & 1 \end{vmatrix} = 3$, $\begin{vmatrix} 5 & 6 & 1 \\ 4 & 6 & 3 \\ 5 & 2 & 6 \end{vmatrix} = 4$, $\begin{vmatrix} 5 & 2 & 1 \\ 4 & 1 & 3 \\ 5 & 1 & 6 \end{vmatrix} = 3$, $\begin{vmatrix} 6 & 2 & 1 \\ 6 & 1 & 3 \\ 2 & 1 & 6 \end{vmatrix} = 4$.

Every square submatrix of A is non-singular. From A we may obtain two MDS codes. These are namely the $[8, 3, -]$ code over $GF(7)$ with the generator matrix $G = (\, I_3 \quad A \,)$ and the $[8, 5, -]$ code over $GF(7)$ with the parity check matrix $H = (\, A \quad I_3 \,)$.

#

For the $[8, 3, -]$ code, the generating function is

$$G = \begin{pmatrix} 1 & 0 & 0 & 3 & 5 & 6 & 2 & 1 \\ 0 & 1 & 0 & 4 & 4 & 6 & 1 & 3 \\ 0 & 0 & 1 & 2 & 5 & 2 & 1 & 6 \end{pmatrix}$$

#

Then,

$$(\,a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5 \quad a_6 \quad a_7 \quad a_8\,) =$$
$$(\,a_1 \quad a_2 \quad a_3\,)\begin{pmatrix} 1 & 0 & 0 & 3 & 5 & 6 & 2 & 1 \\ 0 & 1 & 0 & 4 & 4 & 6 & 1 & 3 \\ 0 & 0 & 1 & 2 & 5 & 2 & 1 & 6 \end{pmatrix}$$

and the encoding functions become

$$\begin{aligned} a_4 &= 3a_1 + 4a_2 + 2a_3 \\ a_5 &= 5a_1 + 4a_2 + 5a_3 \\ a_6 &= 6a_1 + 6a_2 + 2a_3 \\ a_7 &= 2a_1 + a_2 + a_3 \\ a_8 &= a_1 + 3a_2 + 6a_1 \end{aligned}$$

#

The code words are

$C = \{10035621, 01044613, 00125216, 11002534, 10153130, 01162122\}$

#

For the $[8, 5, -]$ code, from the parity check matrix we know that the generating function is

$$G = (\,I_5 \quad -A^T\,) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 4 & 3 & 5 \\ 0 & 1 & 0 & 0 & 0 & 2 & 3 & 2 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 5 \\ 0 & 0 & 0 & 1 & 0 & 5 & 6 & 6 \\ 0 & 0 & 0 & 0 & 1 & 6 & 4 & 1 \end{pmatrix}$$

Then,

$$(\,a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5 \quad a_6 \quad a_7 \quad a_8\,) =$$
$$(\,a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5\,)\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 4 & 3 & 5 \\ 0 & 1 & 0 & 0 & 0 & 2 & 3 & 2 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 5 \\ 0 & 0 & 0 & 1 & 0 & 5 & 6 & 6 \\ 0 & 0 & 0 & 0 & 1 & 6 & 4 & 1 \end{pmatrix}$$

and the encoding functions become

$$\begin{aligned} a_6 &= 4a_1 + 2a_2 + a_3 + 5a_4 + 6a_5 \\ a_7 &= 3a_1 + 3a_2 + a_3 + 6a_4 + 4a_5 \\ a_8 &= 5a_1 + 2a_2 + 5a_3 + 6a_4 + a_5 \end{aligned}$$

#

The code is then

$$C = \left\{ \begin{matrix} 43510000, & 23201000, & 11500100, & 56600010, & 64100001, \\ 66011000, & 54310100, & 22410010, & 30610001, & 34001100, \\ 02101010, & 10301001, & 60400110, & 05600101, & 43000011 \end{matrix} \right\}$$

#

Cryptography

17^{th} February 2006

Definition 101. A *cryptosystem* is a system which modifies a message in such a way that it becomes unintelligible to anyone but the intended recipient. The process used in carrying this out is called *encryption*. A message thus encrypted is called *ciphertext*. The process by which a ciphertext is turned back into plaintext is called *decryption*. The art and science of encrypting messages is called *encryption*, whereas that of decrypting ciphertext without the key is called *cryptanalysis*. Both cryptography and cryptanalysis make up a branch of mathematics called *cryptology*. Let m be a plaintext message, also denoted by p, $e(\cdot)$ the *encryption*, $d(m)$ the *decryption*, c an encrypted string, also known as *cipher*, *ciphertext*, or *cryptogram*, and k a *key*, that is a set of parameters. Then $d(e(m)) = m$ and $c = e(m, k)$. The range of all possible values of the key is called the *keyspace*.

§

Definition 102. There are two kinds of key-based algorithms, namely symmetric and public-key algorithms. *Symmetric* algorithms use the same key for both encryption and decryption. It is also known as *secret-key*, *single-key*, or *one-key* algorithms. There are two kinds of symmetric algorithms, stream and block ciphers. *Stream* algorithms work on a single bit at a time while *block* algorithms work on a group of bits. *Public-key* algorithms use different keys for encryption and decryption. The *encryption key* is called the *public key*, while the *decryption key* the *private key*. Encryption using public key is denoted by $e_k(p) = c$, decryption using the corresponding private key by $d_k(c) = p$. On the other hand, encryption using private key and decryption using public key, as in the case of digital signatures, are denoted respectively as $e_{k_d}(\cdot)$ and $d_{k_e}(\cdot)$.

§

Definition 103. An attempted cryptanalysis is called an *attack*. A successful attack is called a *method*. Assuming the encryption algorithm is known, there are six types of cryptanalysis attack, namely

a. *Cipher-text-only* attack. Here given $c_i = e_k(p_i)$, $i = 1, \ldots, n$, we deduce either p_i, k, or an algorithm a that gives p_{n+1} from $c_{n+1} = e_k(p_{n+1})$, in other words $a : (c = e_k(p)) \mapsto p$.
b. *Known-plaintext* attack. Here given $c_i = e_k(p_i)$ and the corresponding p_i we deduce either k or $a : (c = e_k(p)) \mapsto p$.
c. *Chosen-plaintext* attack. Here choosing p_i we are given $c_i = e_k(p_i)$ and deduce either k or $a : (c = e_k(p)) \mapsto p$.
d. *Adaptive-chosen-plaintext* attack. Here choosing p_i $(c_{j<i})$ the choices of which are based on the results of previous encryption, we are provided with $c_i = e_k(p_i)$ and try to deduce either k or $a : (c = e_k(p)) \mapsto p$.
e. *Chosen-ciphertext* attack. Here choosing c_i we are given the corresponding $p_i = d_k(c_i)$ and try to deduce k.

f. *Chosen-key* attack. In this case you are given the key. So it is not in fact an attack, but rather only a decryption.

§

Definition 104. An algorithm that is unbreakable in practice is said to be *secure*. A secure algorithm can be *unconditionally secure* if there is not enough information to recover the plaintext no matter how much ciphertext one may have, or it can be *computationally secure*, or simply *strong*, if it cannot be broken with available resources. The amount of computing power and time required to recover the encryption key is called the *work factor*.

§

Definition 105. A *substitution cipher* is one in which each letter in the plaintext is replaced by another letter in the ciphertext. There are four types of substution cipher, namely

a. A *simple* substitution cipher. This is the case where the character replacements are one-to-one. In other words, $p^i \overset{one-one}{\longmapsto} c^i$.
b. A *homophonic* substitution cipher. Here the mapping of characters is one-to-many, that is $p^i \overset{one-many}{\longmapsto} c^i$.
c. A *polyalphabetic* substitution cipher. This is when there is a set of simple substitution ciphers for each character mapping, that is to say, $\left\{ p^i \overset{one-one}{\longmapsto} c^i \right\}$.
d. A *polygramme* substitution cipher. This is the case where substitution is done on blocks of characters instead of a single letter. Here $\mathbf{p}^i \overset{one-one}{\longmapsto} \mathbf{c}^i$.

§

Example 34. The *Caesar cipher* is a simple subsitution cipher in which each plaintext character is replaced by the character three to its right modulo 26, that is $c^i \leftarrow \left(p^i + 3\right)$ in $GF(26)$.

Example 35. *ROT13* is a simple encryption programme commonly found on UNIX systems. It has the procedure as shown in Algorithm 6.

given: c^i
if c^i is in $\{a, \ldots, m, A, \ldots, M\}$ **then**
  $c \leftarrow ((c + 13) \bmod 26)$
else
  $c \leftarrow ((c - 13) \bmod 26)$
endif

Definition 106. A *transposition cipher* is one in which the letters in the plaintext remain the same while their order is changed.

§

Example 36. In a simple *columnar transposition cipher* we write the plaintext horizontally on a piece of graph paper of fixed width. The ciphertext is then read off vertically.

Definition 107. A *one-time pad* encryption algorithm is one which uses a non-repeating set of random key letters.

§

Bibliography

Bruce Schneier. *Applied cryptography.* John Wiley & Sons, 1994
Dominic Welsh. *Codes and cryptography.* Oxford, 1988

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

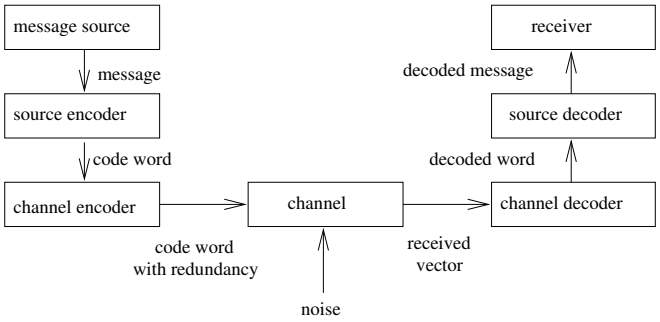


Figure 1 *Encoding and decoding of message. The channel encoder and decoder are there to introduce redundancy which let us detect and correct errors.*

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Criteria for designing channel encoding algorithm and for the construction of the encoder and the decoder are namely fast encoding and decoding of messages, easy transmission of encoded messages, maximum rate of transfer of information, and maximum detection or correction capability.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 1 *Code*

Let $A = \{a_1, \ldots, a_q\}$ be a *code alphabet* of size q, and its elements are the *code symbols*. We call a *q-ary word* of length n over A a sequence $\mathbf{w} = w_1 \cdots w_n$, or equivalently a vector $(w_1, \ldots, w_n)$, where $w_i \in A$ for all i. We call a *q-ary block code* of length n over A a nonempty set C of q-ary words, that is *code words*, all of which is of the same length n. The number of code words C contains is the *size* m of C, consequently $m = |C|$. The *information rate* of the code C is

$$\frac{(\log_q |C|)}{n}$$

We call an $(n, m) - code$ a code of length n and size m.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 1 *Codes*

A code over the code alphabet $\mathbf{F}_2 = \{0, 1\}$ is called a *binary code*, one over $\mathbf{F}_3 = \{0, 1, 2\}$ is called a *ternary code*. The term *quaternary code* refers to a code over either $\mathbf{F}_4 = \{0, 1, 2, 3\}$ or $\mathbf{Z}_4 = \{0, 1, 2, 3\}$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 2 *Communication channel*

A *communication channel* consists of a finite *channel alphabet* $A = \{a_1, \ldots, a_q\}$ together with a set of *forward channel probabilities* $p_{a_{ij}}$, such that for all i

$$\sum_{j=1}^{q} p_{a_{ij}} = 1$$

where $p_{a_{ij}}$ is the conditional probability that a_j is received, given that a_i is sent. If $\mathbf{x}$ is the word received when a word $\mathbf{c}$ was sent, e is the number of places where $\mathbf{x}$ and $\mathbf{c}$ differ, and n the length of each word, then the forward channel probability is

$$p_{\mathbf{cx}} = p^e(1-p)^{n-e}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 3 *Memoryless channel*

Let $\mathbf{c} = c_1 \cdots c_n$ and $\mathbf{x} = x_1 \cdots x_n$ be words of length n. Then a communication channel is said to be *memoryless* if

$$p_{\mathbf{cx}} = \prod_{i=1}^{n} p_{c_i x_i}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 4 *Symmetric channel*

A memory less channel with a channel alphabet of size q is called a *q-ary symmetric channel* if each symbol transmitted has the same probability $p < \frac{1}{2}$ of being received in error, and whenever a wrong symbol is received, each of the $q-1$ possible errors is equally likely.

If $p > \frac{1}{2}$, the channel is known to be *useless*.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 2 *Binary symmetric channel*

The *binary symmetric channel* (BSC) is a memoryless channel having a channel alphabet $\{0,1\}$ and channel probabilities $p_{01} = p_{10} = p$ and $p_{00} = p_{11} = 1-p$.

This probability of a bit error p in a BSC is called the *cross-over probability* of the BSC.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 3 *The most likely word sent*

When a received word is not among the vocabulary of the code, the most likely word sent is the one whose $p_{\mathbf{c}_i\mathbf{x}_i}$ is maximum over all $i = 1, \ldots, m$.

A rule for finding the most likely code word sent in case of an error is called a *decoding rule.*

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 5 *Macimum likelihood decoding*

The *maximum likelihood decoding* is

$$p_{\mathbf{c}_i^x\mathbf{x}} = \max_{\mathbf{c}\in C} p_{\mathbf{cx}}$$

where $\mathbf{x}$ is the word received.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Algorithm 1 *Decoding algorithm*

for all words $\mathbf{x}_i$ received **do**
  if $\mathbf{x}_i$ *is not a valid code word* **then**
    $\mathbf{c}_i^x \leftarrow$*the most likely* $\mathbf{c}_i$ *according to the decoding rule*
  else
    $\mathbf{c}_i^x \leftarrow \mathbf{x}_i$
  endif
endfor

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 4

Two kinds of maximum likelihood decoding are, when it happens that there are more than one word that has the same maximum likelihood, the *complete maximum likelihood decoding* chooses one of them arbitrarily, while the *incomplete maximum likelihood decoding* rejects all of them and asks for a retransmission.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 6 *Hamming distance*

Let $\mathbf{x} = x_1 \cdots x_n$ and $\mathbf{y} = y_1 \cdots y_n$ be words of length n over an alphabet A. Then the *Hamming distance* between $\mathbf{x}$ and $\mathbf{y}$, denoted $\mathrm{d}(\mathbf{x}, \mathbf{y})$, is the number of places where $\mathbf{x}$ and $\mathbf{y}$ are different from each other, and

$$\mathrm{d}(\mathbf{x}, \mathbf{y}) = \mathrm{d}(x_1, y_1) + \ldots + \mathrm{d}(x_n, y_n)$$

where

$$\mathrm{d}(x_i, y_i) = \begin{cases} 0 & \text{if } x_i = y_i \\ 1 & \text{if } x_i \neq y_i \end{cases}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 1 *Hamming distance and the forward channel probability*

The Hamming distance $\mathrm{d}(\mathbf{x}, \mathbf{c}) = i$ corresponds to the forward channel probability

$$p_{\mathbf{cx}} = p^i (1-p)^{n-i}$$

Proof. This is obvious from Definition's 2 and 6. □

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 5

From Definition 6 it follows that

$$0 \leq \mathrm{d}(\mathbf{x}, \mathbf{y}) \leq n$$

$\mathrm{d}(\mathbf{x}, \mathbf{y}) = 0$ if and only if $\mathbf{x} = \mathbf{y}$; and

$$\mathrm{d}(\mathbf{x}, \mathbf{y}) = \mathrm{d}(\mathbf{y}, \mathbf{x})$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 6

Let A be the roman alphabet.

If $\mathbf{x}$ = 'breed', $\mathbf{y}$ = 'bread', and $\mathbf{z}$ = 'break', then

$\mathrm{d}(\mathbf{x}, \mathbf{y}) = \mathrm{d}(\mathbf{y}, \mathbf{z}) = 1$, and $\mathrm{d}(\mathbf{x}, \mathbf{z}) = 2$.

On the other hand if $A = \{0, 1, 2, 3, 4, 5, 6\}$, $\mathbf{p} = 24601$ and $\mathbf{q} = 54321$, then

$$\mathrm{d}(\mathbf{p}, \mathbf{q}) = 3$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 2

Let $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ be words of length n over A. Then the triangular inequality for their mutual Hamming distance holds, that is

$$\mathrm{d}(\mathbf{x},\mathbf{z}) \leq \mathrm{d}(\mathbf{x},\mathbf{y}) + \mathrm{d}(\mathbf{y},\mathbf{z})$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proof. Let $a = \mathrm{d}(\mathbf{x},\mathbf{z})$, $b = \mathrm{d}(\mathbf{x},\mathbf{y})$, and $c = \mathrm{d}(\mathbf{y},\mathbf{z})$. We have $a \geq 0$, $b \geq 0$ and $c \geq 0$.

What this theorem states is obvious when $a = 0$.

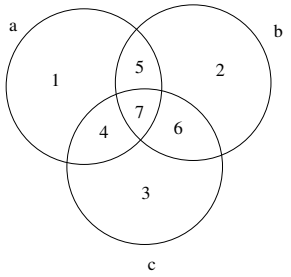
If $a > 0$, then either $b = 0$ or $b > 0$; if the former is the case, that is $b = 0$, then $a = c$ and the theorem is true.

If both $a > 0$ and $b > 0$, then either $c = 0$ or $c > 0$; if $c = 0$, then $a = b$ and the theorem is again true.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

But if $a > 0$, $b > 0$ and $c > 0$, then a, b and c may come from some of the diffences in common, as could be shown in the following Venn diagram.

Figure 2 *Common differences among a, b and c.*



Coding theory, Error and distance, 28^{th} October 2005 –19– From 23^{rd} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Let (x, y) be the differences in common between distances x and y, and similarly (x, y, z) those among x, y and z.

Then from Figure 2 the area 1 is (a); 2, (b); 3, (c); 4, (a, c); 5, (a, b); 6, (b, c); and 7, (a, b, c).

Then, $\mathrm{d}(\mathbf{x}, \mathbf{z})$ arises from the differences $(a) + (c) + (a, b) + (b, c)$,

$\mathrm{d}(\mathbf{x}, \mathbf{y})$ from $(a) + (b) + (a, c) + (b, c)$,

$\mathrm{d}(\mathbf{y}, \mathbf{z})$ from $(b) + (c) + (a, c) + (a, b)$,

and therefore $\mathrm{d}(\mathbf{x}, \mathbf{y}) + \mathrm{d}(\mathbf{y}, \mathbf{z})$ gives $(a) + (b) + (c) + (a, c) + (a, b) + (b, c)$, which is never less than in the case of $\mathrm{d}(\mathbf{x}, \mathbf{y})$ and hence the theorem is again true. This exhausts all the cases and the theory is proved.

Coding theory, Error and distance, 28^{th} October 2005 –20– From 23^{rd} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 7 *Minimum distance decoding*
The *minimum distance-* or *nearest neighbour* decoding rule decodes $\mathbf{x}$ to $\mathbf{c_x}$ if

$$\mathrm{d}(\mathbf{x}, \mathbf{c_x}) = \min_{\mathbf{c} \in C} \mathrm{d}(\mathbf{x}, \mathbf{c})$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 3 *Maximum likelihood- and minimum distance decoding rules*
The maximum likelihood decoding rule and the minimum distance decoding rule is the same for a BSC with cross-over probability

$$p < \frac{1}{2}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proof. From Theorem 1, when $p < \frac{1}{2}$, gives

$$p^0(1-p)^n > \cdots > p^n(1-p)^0$$

Thus the less the distance the more the likelihood, and thus the theorem is proved. □

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 8 *Distance of a code*

Let C be a code containing at least two words. Then, the *minimum distance* or the *distance* of C is

$$\mathrm{d}(C) = \min\{\mathrm{d}(\mathbf{x},\mathbf{y}) | \mathbf{x},\mathbf{y} \in C, \mathbf{x} \neq \mathbf{y}\}$$

A code of length n, size m, and distance d is called

an (n,m,d)-*code.*

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 9 *Error vector*

Let a code word be of length n.

Then, an *error vector* of weight k is a word containing all the k errors occured taking the value of 1 in their corresponding positions with the remaining positions of the word being zero.

An error vector is also called an *error word* or an *error pattern.*

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 10 *Detected and undetected errors*

An error vector is said to be *detected* by a code if $a + e$ is not a code word for any code word a.

If there exists some code word a such that $a + e$ is also a code word, we say that the error vector e goes *undetected.*

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 11 *u-error-detecting code*

Let a received word $\mathbf{x}$ differ from the actual code word sent $\mathbf{c}$ by e errors. Then the corresponding code C is said to be *u-error-detecting* if $\mathbf{x}$ is not a code word whenever $1 \leq e \leq u$.

Moreover, C is *exactly u-error-detecting* if it is *u-error-detecting* but not $(u+1)$*-error-detecting.*

Coding theory, Error and distance, 28^{th} October 2005 –27– From 23^{rd} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 4 *Distance of a u-error-detecting code*

A code C is u-error-detecting if and only if

$$\mathrm{d}(C) \geq u + 1$$

Coding theory, Error and distance, 28^{th} October 2005 –28– From 23^{rd} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proof. Let $\mathbf{c} \in C$.

If $\mathrm{d}(C) \geq u+1$, then $\mathbf{x}$ such that $1 \leq \mathrm{d}(\mathbf{x}, \mathbf{c}) \leq u < \mathrm{d}(C)$ implies that $\mathbf{x} \notin C$, therefore C is u-error-detecting.

On the other hand, if $\mathrm{d}(C) < u+1$, that is $\mathrm{d}(C) \leq u$,

then there exist $\mathbf{x}_1, \mathbf{x}_2 \in C$ such that

$1 \leq \mathrm{d}(C) \leq \mathrm{d}(\mathbf{x}_1, \mathbf{x}_2) \leq u$, then it is possible to send $\mathbf{c}_1 \in C$ and incur errors such that

$1 \leq \mathrm{d}(\mathbf{x}, \mathbf{c}_1) = \mathrm{d}(\mathbf{c}_2, \mathbf{c}_1) \leq u$ and $\mathbf{x} = \mathbf{c}_2$, hence C is not a u-error-detecting code. □

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Corollary 4[1] *Distance of a u-error-detecting code*
A code with distance d is exactly $(d-1)$-error-detecting.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 12 *v-error-correcting code*

Let v be a positive integer and assuming the incomplete decoding rule is used. Then a code C is said to be *v-error-correcting* if the minimum distance decoding can correct for it up to v errors. It is said to be *exactly v-error-correcting* if it is v-error-correcting but not $(v+1)$-error-correcting.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 5 *v-error-correcting code*

A code C is v-error-correcting if and only if

$$\mathrm{d}(C) \geq 2v+1$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proof. Suppose that $\mathrm{d}(C) \geq 2v + 1$.

Let $\mathbf{c} \in C$ be the code word sent, $\mathbf{x}$ the word received, and e errors occurred such that $e \leq v$.

Then $\mathrm{d}(\mathbf{x}, \mathbf{c}) \leq v$, and if C is not to be v-error-correcting there must be some $\mathbf{c}_1, \mathbf{c}_2 \in C$ such that

$\mathrm{d}(\mathbf{x}, \mathbf{c}_1) + \mathrm{d}(\mathbf{x}, \mathbf{c}_2) \leq 2v$.

But since $\mathrm{d}(C) \geq 2v + 1$, which means that $\mathrm{d}(\mathbf{x}, \mathbf{c}_1) + \mathrm{d}(\mathbf{x}, \mathbf{c}_2) \geq 2v + 1$ for all $\mathbf{c}_1, \mathbf{c}_2 \in C$, it follows that C must be v-error-correcting.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Next, suppose that C is v-error-correcting and

$$\mathrm{d}(C) < 2v + 1$$

Then

$$\mathrm{d}(C) \leq 2v$$

that is to say, there exist $\mathbf{c}_1, \mathbf{c}_2 \in C$ such that $\mathrm{d}(\mathbf{c}_1, \mathbf{c}_2) \leq 2v$. This means that there exist

$\mathbf{x}$ such that $\mathrm{d}(\mathbf{x}, \mathbf{c}_1) + \mathrm{d}(\mathbf{x}, \mathbf{c}_2) = \mathrm{d}(\mathbf{c}_1, \mathbf{c}_2) \leq 2v$, hence C is not v-error-correcting. This contradicts what we have supposed earlier, therefore necessarily $\mathrm{d}(C) \geq 2v + 1$. □

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Corollary 5[1] *v-error-correcting code*

A code with distance d is exactly

$\left\lfloor \frac{(d-1)}{2} \right\rfloor$-error-correcting code,

where $\lfloor x \rfloor$ is the greatest integer less than or equal to x.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 13 *probability space and expectation*

A *probability space* is a triple $(\mathrm{S}, \mathrm{B}, \mathrm{P})$ on the domain S, which is a nonempty set called the *sample space*, where (S, B) is a measurable space, B is a Borel field of subsets of S, and P is a measure on S with the property that $\mathrm{P}(\mathrm{S}) = 1$ and, for all disjoint $E_i \in \mathrm{B}$,

$$\mathrm{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathrm{P}(E_i)$$

In other words, P is a nonnegative function defined for all events $E_i \in \mathrm{B}$, and B measurable subsets of S. Further, a *random variable* X is a function mapping S into some set R, called the range of X. For convenience, we shall also use X to represent both the function and its own range, that is X is a function which maps S into X. If S is discrete and f is some real-valued function defined on S, then both X and f(X) are two different random variables, and the expectation of the latter is given by,

$$\mathrm{E}\left[\mathrm{f(X)}\right] = \sum_{x} \mathrm{p}(x)\mathrm{f}(x)$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 14 *conditional probability*

Let $\mathrm{p}(x)$ be the probability that $x \in \mathrm{X}$ occurs, similarly $\mathrm{p}(y)$ that $y \in \mathrm{Y}$ does-, while $\mathrm{p}(x, y)$ that both $x \in \mathrm{X}$ and $y \in \mathrm{Y}$ do occur. Then,

$$\mathrm{p}(x|y) = \frac{\mathrm{p}(x, y)}{\mathrm{p}(y)} \tag{1}$$

and

$$\mathrm{p}(y|x) = \frac{\mathrm{p}(x, y)}{\mathrm{p}(x)} \tag{2}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 15 *Markov chain*

A *Markov chain* is a set of random variable X_t, where $t = 0, 1, \ldots$, such that,

$$\mathrm{P}\left(\mathrm{X}_t = j | \mathrm{X}_0 = i_0, \ldots, \mathrm{X}_{t-1} = i_{t-1}\right) = \mathrm{P}\left(\mathrm{X}_t = j | \mathrm{X}_{t-1} = i_{t-1}\right)$$

In other words, given the present state, the next state is conditionally independent of the past.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 16 *convexity*

A subset $\mathrm{K} \subseteq \mathrm{E}^n$, where E^n is the Euclidean space of n dimensions, is called *convex* if the line segment joining any two points in K is contained in K. Let the two points be x_1 and x_2, then the line segment joining them together is

$$x = tx_1 + (1 - t)x_2$$

where $0 \leq t \leq 1$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 17 *convex hull*

A point x is said to be a *convex combination* of points $x_1, \ldots, x_m$ if there exist nonnegative scalars $\alpha_1, \ldots, \alpha_m$ such that

$$\sum \alpha_i = 1$$

and

$$\sum \alpha_i x_i = x$$

The set of all convex combinations of x_i, $i = 1, \ldots, m$, is called the *convex hull* of $\{x_i\}$.

Coding theory, Entropy and mutual information, 4^{th} November 2005 –5– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 18 *convex cup and cap*

Let f be a real-valued function, and let K be a convex subset of the domain of f. Then f is said to be *convex cup* if, for every $x_1, x_2 \in$ K and $0 \leq t \leq 1$,

$$\mathrm{f}(tx_1 + (1-t)x_2) \leq t\mathrm{f}(x_1) + (1-t)\mathrm{f}(x_2) \tag{3}$$

It is said to be *strictly convex cup* if strict inequality holds in Equation 3 whenever $x + 1 \neq x_2$. Similarly, f is said to be *convex cap* if,

$$\mathrm{f}(tx_1 + (1-t)x_2) \geq t\mathrm{f}(x_1) + (1-t)\mathrm{f}(x_2) \tag{4}$$

that is to say, if $-f$ is convex cup. It is said to be *strictly convex cap* if strict inequality holds in Equation 4 whenever $x_1 \neq x_2$. Convex cap is also known as *concave*. Geometrically speaking, f is convex cup if and only if all its chords lie above or on the graph of f, and f is concave if and only if all its chords lie below or on the graph of the same.

Coding theory, Entropy and mutual information, 4^{th} November 2005 –6– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 19 *Jensen's inequality*

Let K be some interval in E^1, and let $\mathrm{F}(x)$ be a probability distribution concentrated on K such that

$$\mathrm{P}(\mathrm{X} \leq x) = \mathrm{F}(x)$$

Then, if the expectation E(X) exists, and if $\mathrm{f}(x)$ is a convex cup function, then,

$$\mathrm{E}(\mathrm{f}(\mathrm{X})) \geq \mathrm{f}(\mathrm{E}(\mathrm{X})) \tag{5}$$

If f is strictly convex cup, then strict inequality holds in Equation 5. Similarly, if f is convex cap, then,

$$\mathrm{E}(\mathrm{f}(\mathrm{X})) \leq \mathrm{f}(\mathrm{E}(\mathrm{X}) \tag{6}$$

If f is strictly convex cap, then strict inequality holds in Equation 6.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 7 *Jensen's inequality in geometrical terms*

Suppose that in Definition 19 there is a mass distribution placed on the graph of f, then Equation 5 says that the overall centre of mass will lie above or on the graph, while Equation 6 says that it will lie below it.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Axiom 1 *Entropy*

If the events are all equally likely, then the uncertainty function

$$\mathrm{H}\left(\frac{1}{m},\ldots,\frac{1}{m}\right)$$

is monotonously increasing with m.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Axiom 2 *Entropy*

If $\{E_1^1,\ldots,E_m^1\}$ and $\{E_1^2,\ldots,E_n^2\}$ are statistically independent sets of equally likely disjoint events, then the uncertainty of the sets of events

$$\{E_i \cap E_j; i=1,\ldots,m; j=1,\ldots,n\}$$

is

$$\mathrm{H}\left(\frac{1}{mn},\ldots,\frac{1}{mn}\right)=\mathrm{H}\left(\frac{1}{m},\ldots,\frac{1}{m}\right)+\mathrm{H}\left(\frac{1}{n},\ldots,\frac{1}{n}\right)$$

That is to say,

$$\mathrm{h}(mn)=\mathrm{h}(m)+\mathrm{h}(n)$$

where

$$\mathrm{h}(m)=\mathrm{H}\left(\frac{1}{m},\ldots,\frac{1}{m}\right)$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 20 *Entropy*

Let the set of m possible disjoint events be $\mathrm{E} = \{E_1, \ldots, E_m\}$. We call an *apriori probability* of E_i, $\mathrm{p}(E_i)$, where $1 \leq i \leq m$ and

$$\sum_{i=1}^{m} \mathrm{p}(E_i) = 1$$

The *uncertainty function* or the *entropy function*,

$$\mathrm{H}(\mathrm{p}(1), \ldots, \mathrm{p}(m))$$

obeys Axiom's 1 and 2.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 6 *Entropy for equally likely events*

The entropy of a set of m equally likely events is

$$\mathrm{h}(m) = \lambda \log_c m$$

where λ is a positive constant and $c > 1$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proof. Proving Theorem 6 amounts to proving that Axiom's 1 and 2 are satisfied if and only if $\mathrm{h}(m) = \lambda \log_c m$. The two axioms say that $\mathrm{h}(m)$ is monotonously increasing in m and

$$\mathrm{h}(mn) = \mathrm{h}(m) + \mathrm{h}(n) \tag{7}$$

According to Equation 7, if $m = n = 1$, then $\mathrm{h}(1) = \mathrm{h}(1) + \mathrm{h}(1)$, which implies that $\mathrm{h}(1) = 0$. From this together with both axioms above, $\mathrm{h}(m) = \lambda \log_c m$ is sufficient as a solution.
Next, we must prove that this solution is necessarily the only solution. Let a, b and c be positive integers, and $a, b, c > 1$. Then there exists a unique integer d such that

$$c^d \leq a^b < c^{d+1} \tag{8}$$

Coding theory, Entropy and mutual information, 4^{th} November 2005 –13– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

From Equation 8 it follows that,

$$d \log c \leq b \log a < (d+1) \log c$$

and therefore,

$$\frac{d}{b} \leq \frac{\log a}{\log c} < \frac{d+1}{b} \tag{9}$$

Since $\mathrm{h}(m)$ is monotonously increasing, from Equation 8 we have,

$$\mathrm{h}(c^d) \leq \mathrm{h}(a^b) < \mathrm{h}(c^{d+1})$$

Then from Equation 7, $d\mathrm{h}(c) \leq b\mathrm{h}(a) < (d+1)\mathrm{h}(c)$. And since $\mathrm{h}(m)$ is monotonously increasing,

$$\frac{d}{b} \leq \frac{\mathrm{h}(a)}{\mathrm{h}(c)} < \frac{d+1}{b} \tag{10}$$

Coding theory, Entropy and mutual information, 4^{th} November 2005 –14– From 25^{th} October 2005 , as of 14^{th} January, 2007

From Equation's 9 and 10 it follows that,

$$\left|\frac{\log a}{\log c} - \frac{\mathrm{h}(a)}{\mathrm{h}(c)}\right| < \frac{1}{b}$$

And, since b is arbitrary positive integer,

$$\frac{\mathrm{h}(a)}{\mathrm{h}(c)} = \frac{\log a}{\log c}$$

$$\frac{\mathrm{h}(a)}{\log a} = \frac{\mathrm{h}(c)}{\log c}$$

Since a and c are arbitrary,

$$\frac{\mathrm{h}(a)}{\log a} = \lambda = \frac{\mathrm{h}(c)}{\log c}$$

Therefore, necessarily $\mathrm{h}(m) = \lambda \log_c m$ is the only solution. ¶

Axiom 3 *grouping*

The total uncertainty of events does not depend on the method of indication.

§

Axiom 4 *continuity of entropy*

The uncertainty measure is a continuous function with regard to the probabilities within it.

§

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 8 *grouping events*

Let a set E of m disjoint events be $\{E_1, \ldots, E_m\}$. Let j_i, $i = 0, \ldots, n$, be integers and $0 = j_0 \leq j_1 < j_2 \cdots < j_n = m$, and E be divided into n sets of events, namely,

$$\begin{aligned} \mathrm{G}_1 &= \{E_1, \ldots, E_{j_1}\} \\ \mathrm{G}_2 &= \{E_{j_1+1}, \ldots, E_{j_2}\} \\ &\vdots \\ \mathrm{G}_n &= \{E_{j_{n-1}+1}, \ldots, E_m\} \end{aligned}$$

If we indicate firstly the group, and then the event within that group, then the uncertainty becomes,

$$\begin{aligned} \mathrm{H}(\mathrm{p}(E_1), \ldots, \mathrm{p}(E_m)) &= \mathrm{H}(\mathrm{p}(\mathrm{G}_1), \ldots, \mathrm{p}(\mathrm{G}_n)) \\ &\quad + \sum_{i=1}^{n} \mathrm{p}(\mathrm{G}_i)\mathrm{H}(\mathrm{p}(E_{j_{i-1}+1}|\mathrm{G}_i), \ldots, \mathrm{p}(\mathrm{E}_{j_i}|\mathrm{G}_i)) \end{aligned} \tag{11}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

The grouping axiom, Axiom 3, lets us express the uncertainty when all the event probabilities are rational. By grouping equally likely events together and then consider each of the groups as a single event, it gives us the ability to deal with events which are not equally likely. Example 9 gives an example how this is done. Then Axiom 4 extends Axiom 3 to cover also irrational probabilities, and Equation 12 is the result.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 9 *entropy of groups of events*

As in Example 8, let a set of disjoint events be $\mathrm{E} = \{E_1, \ldots, E_m\}$, and let $\mathrm{p}(E_i) = \frac{1}{m}$, $i = 1, \ldots, m$. Also, let the groups of events $\mathrm{G}_1, \ldots, \mathrm{G}_n$ be defined the same way therein. Let n_k be the number of events in G_k. Then $n_k = j_k - j_{k-1}$ and $\mathrm{p}(\mathrm{G}_k) = \frac{n_k}{m}$, for $k = 1, \ldots, n$, and also $\mathrm{p}(E_i|\mathrm{G}_k) = \frac{1}{n_k}$, for $j_{k-1} < i \leq j_k$. Then Equation 11 yields, $\mathrm{h}(m) = \mathrm{H}(\mathrm{p}(\mathrm{G}_1), \ldots, \mathrm{p}(\mathrm{G}_n)) + \sum_{i=1}^{n} \mathrm{p}(\mathrm{G}_i)\mathrm{h}(n_i)$. And since from Theorem 6, $\mathrm{h}(m) = \lambda \log_c m$, we have,

$$\mathrm{H}(\mathrm{p}(\mathrm{G}_1), \ldots, \mathrm{p}(\mathrm{G}_n)) = -\sum_{i=1}^{n} \mathrm{p}(\mathrm{G}_i)(\mathrm{h}(n_i) - \mathrm{h}(m))$$

$$= -\sum_{i=1}^{n} \mathrm{p}(\mathrm{G}_i) \left(\lambda \log \frac{n_i}{m}\right) = -\lambda \left(\sum_{i=1}^{n} \mathrm{p}(\mathrm{G}_i) \log \mathrm{p}(\mathrm{G}_i)\right) \tag{12}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 10 *scaling the entropy*

From

$$\mathrm{h}(m) = \lambda \log_c m$$

if we let

$$\lambda = \log_b c$$

then

$$\mathrm{h}(m) = \log_b m$$

In other words, the scale factor λ can be absorbed in the base of the logarithm.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 7 *entropy*

Let $\{p_1, \ldots, p_m\}$ be a set of probabilities such that

$$\sum_{i=1}^{m} p_i = 1$$

Then, †

$$\mathrm{H}(p_1, \ldots, p_m) = -\sum_{i=1}^{m} p_i \log p_i \tag{13}$$

Proof. This is the results from Example's 8 and 9, and the scale factor λ disappears in a manner similar to that shown by Example 10. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 11 *units of entropy*

If the base of the logarithm in Equation 13 is 2, the unit of the entropy is *bit*. On the other hand if this base is e, that is to say, if we use natural logarithms, then the uncertainty has the unit of *nat*. From this, one may see that one nat is equal to $\log_2 e$ bits, which is approximately 1.443 bits. The term *bit* comes from *binary digit*, the term *nat* from *natural digit*.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 21 *conditional entropy*

The *conditional entropy* of X, given some $y \in \mathrm{Y}$, is,

$$\mathrm{H}(\mathrm{X}|y) = -\sum_{x} \mathrm{p}(x|y) \log \mathrm{p}(x|y) \tag{14}$$

Then the conditional entropy H(X|Y) is the expectation, or average value, of H(X|y) over the range Y. In other words,

$$\mathrm{H}(\mathrm{X}|\mathrm{Y}) = \sum_{y} \mathrm{p}(y)\mathrm{H}(\mathrm{X}|y) \tag{15}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 8 *conditional entropy*

The conditional entropy is,

$$\mathrm{H}(\mathrm{X}|\mathrm{Y}) = -\sum_{x,y} \mathrm{p}(x,y) \log \mathrm{p}(x|y)$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proof. Putting the equation of conditional entropy when y is given, Equation 14, into the overall conditional entropy equation, Equation 15, we get,

$$\begin{aligned} \mathrm{H(X|Y)} &= \sum_y \mathrm{p}(y)\mathrm{H(X}|y) \\ &= -\sum_y \mathrm{p}(y) \sum_x \mathrm{p}(x|y) \log \mathrm{p}(x|y) \end{aligned}$$

Then from Equation 1 of Definition 14, $\mathrm{p}(y)\mathrm{p}(x|y) = \mathrm{p}(x,y)$, and so,

$$\mathrm{H(X|Y)} = -\sum_{x,y} \mathrm{p}(x,y) \log \mathrm{p}(x|y)$$

¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 9 *conditional entropy*

Let X, Y and Z be discrete random variables. For each $z \in \mathrm{Z}$, let

$$\mathrm{E}(z) = \sum_{x,y} \mathrm{p}(y)\mathrm{p}(z|x,y)$$

Then,

$$\mathrm{H(X|Y)} \leq \mathrm{H(Z)} + \mathrm{E}(\log \mathrm{E})$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proof.

$$\begin{aligned} \mathrm{H(X|Y)} &= -\mathrm{E}\left[\log \mathrm{p}(x|y)\right] \\ &= -\sum_{x,y,z} \mathrm{p}(x,y,z) \log \mathrm{p}(x|y) \\ &= -\sum_{z} \mathrm{p}(z) \sum_{x,y} \frac{\mathrm{p}(x,y,z)}{\mathrm{p}(z)} \log \mathrm{p}(x|y) \end{aligned}$$

Because

$$\frac{\mathrm{p}(x,y,z)}{\mathrm{p}(z)} = \mathrm{p}(x,y|z)$$

is a probability distribution, that is a convex cap function, we may apply Equation 6, namely Jensen's inequality for convex cap, from Definition 19.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Hence,

$$\begin{aligned} \mathrm{H(X|Y)} &\leq \sum_{z} \mathrm{p}(z) \log \left(\frac{1}{\mathrm{p}(z)} \sum_{x,y} \frac{\mathrm{p}(x,y,z)}{\mathrm{p}(x|y)} \right) \\ &= \sum_{z} \mathrm{p}(z) \log \frac{1}{\mathrm{p}(z)} + \sum_{z} \mathrm{p}(z) \log \sum_{x,y} \frac{\mathrm{p}(x,y,z)}{\mathrm{p}(x|y)} \end{aligned}$$

But,

$$\frac{\mathrm{p}(x,y,z)}{\mathrm{p}(x|y)} = \frac{\mathrm{p}(x,y,z)\mathrm{p}(y)}{\mathrm{p}(x,y)} = \mathrm{p}(y)\mathrm{p}(z|x,y)$$

hence the statement above is proved. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Corollary 9[1] *Fano's inequality*

Let X and Y be random variables each of which takes values in the set $\{x_1, \ldots, x_r\}$. Let

$$\mathrm{P}_e = \mathrm{P}(\mathrm{X} \neq \mathrm{Y})$$

Then,

$$\mathrm{H}(\mathrm{X}|\mathrm{Y}) \leq \mathrm{H}(\mathrm{P}_e) + \mathrm{P}_e \log(r-1)$$

Proof. From Theorem 9, let $\mathrm{Z} = 0$ if $\mathrm{X} = \mathrm{Y}$, and let $\mathrm{Z} = 1$ if $\mathrm{X} \neq \mathrm{Y}$. Then $\mathrm{E}(0) = 1$ and $\mathrm{E}(1) = r - 1$. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 10 *maximum entropy*

The maximum uncertainty occurs when the events are equiprobable.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proof. Since,

$$\mathrm{H}\left(\frac{1}{m},\ldots,\frac{1}{m}\right) - \mathrm{H}(p_1,\ldots,p_m) = \log_b m + \sum_{i=1}^{m} p_i \log_b p_i$$
$$= \log_b e \sum_{i=1}^{m} p_i \ln m p_i \geq \log_b e \sum_{i=1}^{m} p_i \left(1 - \frac{1}{m p_i}\right) = 0$$

it being the case that

$$\ln \frac{1}{x} \geq 1 - x$$

Therefore $\mathrm{H}(p_1,\ldots,p_m)$ is maximised when

$$p_i = \frac{1}{m}$$

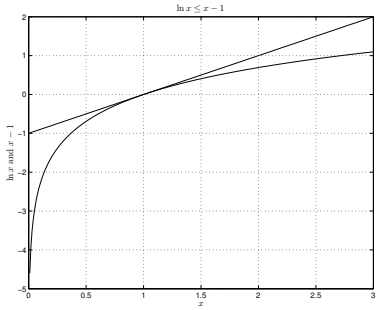
for all $i = 1,\ldots,m$. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 12 *inequality for a bound on* $\ln x$

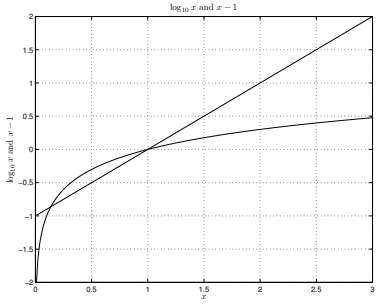
Figure 3 shows that $\ln x \leq x - 1$, while Figure 4 shows that such inequality does not exist when the logarithm in question is of base 10.

Figure 3 *Plots of* $\ln x$ *and* $x - 1$*, which show that* $\ln x \leq x - 1$.



Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Figure 4 *Graphs of* $y = \log x$ *and* $y = x - 1$*, which show that the latter is no bound for the values of the former.*



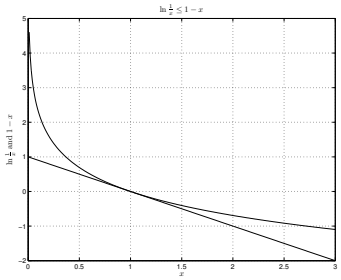
Coding theory, Entropy and mutual information, 4^{th} November 2005 –33– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 13 *inequality for a bound on* $\ln \frac{1}{x}$

Figure 5 confirms for us how $\ln \frac{1}{x} \geq 1 - x$, whereas Figure 6 tells us that this is the case for $\log \frac{1}{x}$.

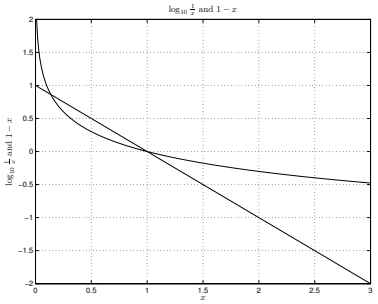
Figure 5 *Plots showing* $\ln \frac{1}{x}$ *and* $1 - x$*, which show that* $\ln \frac{1}{x} \geq 1 - x$.



Coding theory, Entropy and mutual information, 4^{th} November 2005 –34– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Figure 6 *Graphs showing* $y = \log \frac{1}{x}$ *and* $y = 1 - x$*, from which it is clear the latter gives no bounds for the former.*



Coding theory, Entropy and mutual information, 4^{th} November 2005 –35– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 14 *two events*

Consider two events with probabilities p and $1 - p$. The entropy function is then,

$$\mathrm{H}(p, 1-p) = -p \log p - (1-p) \log(1-p)$$

Whenever the occurrence of either event become certainty, the entropy function would become zero. Mathematically we see that

$$\lim_{p \to 0} p \log p = 0$$

and

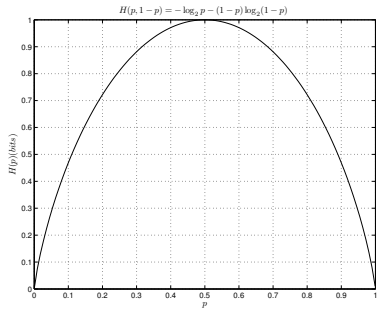
$$\lim_{p \to 1} p \log p = 0$$

Figure 7 shows a plot of the values of the entropy function for two events. Base-2 logarithm is used here.

Coding theory, Entropy and mutual information, 4^{th} November 2005 –36– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Figure 7 *The entropy function of two events with probabilities p and $1-p$.*



Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 22 *mutual information*

The *mutual information* is

$$\mathrm{I(X;Y)} = \mathrm{H(X)} - \mathrm{H(X|Y)}$$

It represents the information provided about X by Y.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 15 *mutual information*

Alternatively, the mutual information may take the following form, *cf* Definition 14,

$$\mathrm{I(X;Y)} = \sum_{x,y} \mathrm{p}(x,y) \log \frac{\mathrm{p}(x|y)}{\mathrm{p}(x)}$$

$$= \sum_{x,y} \mathrm{p}(x,y) \log \frac{\mathrm{p}(x,y)}{\mathrm{p}(x)\mathrm{p}(y)} = \sum_{x,y} \mathrm{p}(x,y) \log \frac{\mathrm{p}(y|x)}{\mathrm{p}(y)}$$

That is to say, $\mathrm{I(X;Y)}$ is the average taken over the X, Y sample space of the random variable $\mathrm{I}(x;y)$ such that,

$$\mathrm{I}(x;y) = \log \frac{\mathrm{p}(x|y)}{\mathrm{p}(x)} = \log \frac{\mathrm{p}(x,y)}{\mathrm{p}(x)\mathrm{p}(y)} = \log \frac{\mathrm{p}(y|x)}{\mathrm{p}(y)}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 11 *mutual information*

For any discrete random variables X and Y,

$$\mathrm{I(X;Y)} \geq 0$$

Moreover,

$$\mathrm{I(X;Y)} = 0$$

if and only if X and Y are independent.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proof. From one of our formulae for the mutual information and from Jensen's inequality,

$$\begin{aligned} \mathrm{I(X;Y)} &= -\sum_{x,y} \log \frac{\mathrm{p}(x)\mathrm{p}(y)}{\mathrm{p}(x,y)} \\ &\geq \log \sum_{x,y} \mathrm{p}(x)\mathrm{p}(y) = \log 1 = 0 \end{aligned}$$

Furthermore, the equality sign holds if and only if

$$\mathrm{p}(x)\mathrm{p}(y) = \mathrm{p}(x,y)$$

for all x and y, that is to say, when X and Y are independent of each other. ¶

Coding theory, Entropy and mutual information, 4^{th} November 2005 –41– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 16 *mutual information in various forms*

From our formulae of the mutual information, we may see that,

$$\mathrm{I(X;Y)} = \mathrm{I(Y;X)}$$

and

$$\mathrm{I(X;Y)} = \mathrm{H(Y)} - \mathrm{H(Y|X)}$$

Also,

$$\mathrm{I(X;Y)} = \sum_{x,y} \mathrm{p}(x,y) \log \frac{1}{\mathrm{p}(x,y)}$$

Coding theory, Entropy and mutual information, 4^{th} November 2005 –42– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 23 *mutual information for three random variables*
Let X, Y and Z be three random variables. Then the mutual information I(X, Y; Z) is given by,

$$\mathrm{I(X,Y;Z)} = \mathrm{E}\left(\log \frac{\mathrm{p}(z|x,y)}{\mathrm{p}(z)}\right) = \sum_{x,y,z} \mathrm{p}(x,y,z) \log \frac{\mathrm{p}(z|x,y)}{\mathrm{p}(z)}$$

This mutual information is the amount of information X and Y provide about Z.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 12 *mutual information for three random variables*
Let X, Y and Z be three random variables. Then we have

$$\mathrm{I(X,Y;Z)} \geq \mathrm{I(Y;Z)}$$

where the equality holds if and only if

$$\mathrm{p}(z|x,y) = \mathrm{p}(z|y)$$

for all (x, y, z) such that

$$\mathrm{p}(x,y,z) > 0$$

Proof.

$$\mathrm{I(Y;Z)} - \mathrm{I(X,Y;Z)} = \mathrm{E}\left(\log \frac{\mathrm{p}(z|y)}{\mathrm{p}(z)} - \log \frac{\mathrm{p}(z|x,y)}{\mathrm{p}(z)}\right)$$

$$= \mathrm{E}\left(\log \frac{\mathrm{p}(z|y)}{\mathrm{p}(z|x,y)}\right) = \sum_{x,y,z} \mathrm{p}(x,y,z) \log \frac{\mathrm{p}(z|y)}{\mathrm{p}(z|x,y)}$$

Then using Jensen's inequality, we have,

$$\mathrm{I(Y;Z)} - \mathrm{I(X,Y;Z)} \leq \log \sum_{x,y,z} \mathrm{p}(x,y,z) \frac{\mathrm{p}(z|y)}{\mathrm{p}(z|x,y)}$$

$$= \log \sum_{x,y,z} \mathrm{p}(x,y)\mathrm{p}(z|y) = \log 1 = 0$$

¶

Theorem 13 *Markov's chain*

Let $(\mathrm{X}, \mathrm{Y}, \mathrm{Z})$ be a Markov chain. Then,

$$\mathrm{I(X;Z)} \leq \begin{cases} \mathrm{I(X;Y)} \\ \mathrm{I(Y;Z)} \end{cases}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proof. From Theorem 12,

$$I(X;Z) \leq I(X,Y;Z)$$

Because (X,Y,Z) is a Markov chain,

$$I(X,Y;Z) = I(Y;Z)$$

Therefore

$$I(X;Z) \leq I(Y;Z)$$

Next, since (X,Y,Z) is a Markov chain, (Z,Y,X) is also a Markov chain. Hence

$$I(X;Z) \leq I(X;Y)$$

¶

Coding theory, Entropy and mutual information, 4^{th} November 2005 –47– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 24 *group*

A *group* is a non-empty set G together with an operation, called *multiplication*, which associates with each ordered pair x, y of elements in G a third element, their *product*, in G such that,

1. multiplication is *associative*;
2. there exists an *identity element* e in G; and
3. for each element x in G there exists an *inverse* of x.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

In other words, for x and y in G there exists xy in G such that,

1. for any x, y and z in G, $x(yz) = (xy)z$;
2. there exists e in G such that $xe = ex = x$; and
3. to each x in G there corresponds x^{-1} in G such that $xx^{-1} = x^{-1}x = e$.

A group is called *Abelian* or *commutative group* if

$$xy = yx$$

for all elements x and y in G. The group G is called a *finite group* if it consists of a finite number of elements, otherwise it is called an *infinite group*. This number of elements of G is called its *order*.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 14 *uniqueness of an identity*
Both the identity e and the inverse x^{-1} of a group G are unique.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proof. Suppose e^90 is another element in G such that

$$xe^90 = e^90x = x$$

for every x in G, then

$$e^90 = e^90e = e$$

hence the identity element is unique. Suppose for every x in G, that x^90 be another element in G such that

$$xx^90 = x^90x = e$$

then,

$$x^90 = x^90e = x^90(xx^{-1}) = (x^90x)x^{-1} = ex^{-1} = x^{-1}$$

hence the inverse element of G is unique.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 25 *ring*

A *ring* is an additive Abelian group R which is closed under a second operation, called *multiplication*, in such a manner that,

1. multiplication is *associative*; and
2. multiplication is *distributive*.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

That is to say, if x, y and z are any three elements in R, then,

1. $x(yz) = (xy)z$; and
2. $x(y + z) = xy + xz$ and $(x + y)z = xz + yz$.

A ring is called a *commutative ring* if

$$xy = yx$$

for all elements x and y in R. If a ring R has a non-zero element 1 with such a property that

$$x1 = 1x = x$$

for every x, then 1 is called an *identity element*, and R is said to be a *ring with identity*.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 26 *regular- and singular elements*

Let x be an element of R, a ring with identity. Then x is said to be *regular* if its inverse x^{-1} exists, otherwise it is said to be *singular*. Regular elements are also called *invertible-* or *non-singular* elements. Furthermore, R is called a *division ring* if all its non-zero elements are regular.

§

Definition 27 *field*

A *field* is a commutative division ring.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 17 *field*

A field, then, is a non-empty set F together with two operations on its elements, namely addition and multiplication, such that for all a, b and c in F, under addition, F is closed, commutative, associative, has a unique identity, has for each of its elements a unique inverse; and under multiplication, F is closed, commutative, associative, has a unique identity, has for each of its elements a unique inverse. Furturemore, F is also distributive.
These properties of field are inherited from the latter's progenitors, since the field is defined by the division ring which itself is defined by the ring which itself is defined by the group.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Table 1 shows the sources from which each of the properties of the field is defined.

operator	*property*	*defining definition*
addition	closed	group
	commutative	Abelian group
	associative	group
	identity	group
	inverse	group
multiplication	closed	ring
	commutative	commutative ring
	associative	ring
	identity	ring with identity
	inverse	division ring

Table 1 *The various sources at the places of which the various properties of the field are defined.*

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 15 *negative*

Consider any two elements a and b in a field F, we have

$$(-1) \cdot a = -a$$

Proof. Since,

$$(-1) \cdot a + a = (-1) \cdot a + a \cdot 1 = ((-1) + 1) \cdot a = 0 \cdot a = 0$$

and since $a + (-a) = 0$, therefore $(-1) \cdot a = -a$. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 16 *zeros*

Let a and b be any two elements in a field F. Then

$$ab = 0 \,\text{implies}\, a = 0 \,\text{or}\, b = 0$$

Proof. If $a \neq 0$, then,

$$0 = a^{-1} \cdot 0 = a^{-1}(ab) = (a^{-1}a)b = 1 \cdot b = b \cdot 1 = b$$

And since a and b are arbitrary, and since $ab = ba$, our statement above is proved. ¶

Coding theory, Group, field and finite field, 11^{th} November 2005 –11– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 28 *modulo*

Let a, b and m be integers, and let $m > 1$. Then a is said to be *congruent to* b *modulo* m, in other words,

$$a \equiv b (\text{mod}\, m)$$

if $m|(a - b)$, that is to say, m divides $a - b$. The number m is called the *modulus*, and b is called the *residue* of $a (\text{mod}\, m)$. Sometimes b is also called the *principal remainder* of a divided by m, and denoted by

$$(a (\text{mod}\, m))$$

A residue is said to be *common* if $0 \leq b < m$.

Coding theory, Group, field and finite field, 11^{th} November 2005 –12– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 17 *congruence*

Any integer a is congruent to exactly one of 0, 1, ..., $m-1$ modulo m.

Proof. Let a and m be integers, and let $m > 1$. Then there exists a unique k such that $a = mk + b$, where $0 \leq b \leq m-1$. Therefore b is uniquely determined by m and a.

To prove that b is unique, suppose there exist $a = mk_1 + b_1$ and $a = mk_2 + b_2$, where $0 \leq b_1 \leq m-1$ and $0 \leq b_2 \leq m-1$, such that $b_1 \neq b_2$. Then, $a - mk_1 \neq a - m_k 2$, and since $m > 1$, therefore $k_1 \neq k_2$. Since k_1 and k_2 are arbitrary, let $k_1 > k_2$ and let $k_1 = k_2 + n$. Then,

$$mk_2 + b_2 = a = m(k_2 + n) + b_1 = mk_2 + b_1 + mn$$

and since $b_1 \geq 0$, $m \geq 0$ and $n > 0$, we have $b_2 \geq m$, which contradicts what we have said earlier, that is $b_2 \leq m - 1$. So, necessarily $b_1 = b_2$. ¶

Coding theory, Group, field and finite field, 11^{th} November 2005 –13– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 18 *properties of modulo*

Let a, b and m are integers, and let $m > 1$. Then the following properties hold for congruence.

a. $a \equiv b (\mathrm{mod}\, 0)$ implies $a = b$
b. either $a \equiv b (\mathrm{mod}\, m)$ or $a \not\equiv b (\mathrm{mod}\, m)$
c. $a \equiv a (\mathrm{mod}\, m)$
d. $a \equiv b (\mathrm{mod}\, m)$ implies $b \equiv a (\mathrm{mod}\, m)$
e. if $a \equiv b (\mathrm{mod}\, m)$ and $b \equiv c (\mathrm{mod}\, m)$, then $a \equiv c (\mathrm{mod}\, m)$

Let $a \equiv b (\mathrm{mod}\, m)$ and $c \equiv d (\mathrm{mod}\, m)$. Then,

f. $a + c \equiv b + d (\mathrm{mod}\, m)$
g. $a - c \equiv b - d (\mathrm{mod}\, m)$
h. $ac \equiv bd (\mathrm{mod}\, m)$

Coding theory, Group, field and finite field, 11^{th} November 2005 –14– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Further, let k and n be integers. Then,

i. if $a \equiv b (\mathrm{mod}\, m)$, then $ka \equiv kb (\mathrm{mod}\, m)$
j. if $a \equiv b (\mathrm{mod}\, m)$, then $a^n \equiv b^n (\mathrm{mod}\, m)$
k. if $a \equiv b (\mathrm{mod}\, m_1)$ and $a \equiv b (\mathrm{mod}\, m_2)$, then,

$$a \equiv b \,(\mathrm{mod}\, \mathrm{lcm}(m_1, m_2))$$

where $\mathrm{lcm}(x, y)$ is the least common multiple of x and y, that is the smallest z such that there exist positive integers p and q by which

$$px = qy = z$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

l. if $a^k \equiv b^k (\mathrm{mod}\, m)$, then,

$$a \equiv b \left(\mathrm{mod}\, \frac{m}{\gcd(k, m)}\right)$$

From above properties, it follows that,

m. if $a \equiv b (\mathrm{mod}\, m)$, then

$$P(a) \equiv P(b) (\mathrm{mod}\, m)$$

where $P(x)$ is a polynomial.

Properties (a) is called *equivalence*, (b) *determination*, (c) *reflexive*, (d) *symmetry*, and (e) *transition*.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 29 *set of integer modulo m*

We denote by $\mathbf{Z}_m$ or $\mathbf{Z}/(m)$ the set $\{0, \ldots, m-1\}$, where $m > 1$, and define the addition and multiplication on it as,

$$a \oplus b = (a + b (\mathrm{mod}\, m))$$

and

$$a \odot b = (ab (\mathrm{mod}\, m))$$

respectively, and these may be denoted as $a + b$ and respectively ab for simplicity.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 18 *ring of integer modulo m*

The set $\mathbf{Z}$ together with addition and multiplication introduced in Definition 29 form a ring.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 19 *condition when* $\mathbf{Z}_m$ *is a field*

The ring $\mathbf{Z}_m$ is a field if and only if m is prime.

Proof. First we prove that m being prime implies that $\mathbf{Z}_m$ is a field. Let m be a prime. Then any $a \neq 0$ in $\mathbf{Z}_m$, in other words $0 < a < m$, is prime relative to m. Therefore, there exist two integers u and v, where $0 \leq u \leq m-1$, such that $ua + vm = 1$, which means that $ua \equiv 1 (\mathrm{mod}\, m)$. Hence $u = a^{-1}$, and since this applies for every a in $\mathbf{Z}_m$, it follows that $\mathbf{Z}_m$ is a field.

Next we will prove that if m is not a prime, then $\mathbf{Z}_m$ is no field. Suppose that m is not a prime. Then $m = ab$ for some a and b, where $1 < a < m$ and $1 < b < m$. But $ab = 0$ is in $\mathbf{Z}_m$, and therefore $a = 0$ and $b = 0$. This contradicts the values of a and b given above, thus $\mathbf{Z}_m$ is no field. ¶

Coding theory, Group, field and finite field, 11^{th} November 2005 –19– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 30 *notation for rings*

We denote by na the element

$$\sum_{i=1}^{n} a$$

for any element a in a ring R and an integer $n \geq 1$.

§

Definition 31 *characteristic of a field*

Let F be a field. Then the *characteristic* of F is the least positive integer p such that $p \cdot 1 = 0$, where 1 is the multiplicative identity of F. Where no such p exists, this characteristic is defined to be zero.
By F^* we mean $F \backslash \{0\}$.

Coding theory, Group, field and finite field, 11^{th} November 2005 –20– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 20 *characteristic of a field*

The characteristics of a field is either zero or a prime number.

Proof. Consider a field F. Since $1 \cdot 1 = 1 \neq 0$, therefore 1 is not the characteristic of F. Let the characteristic be $p = mn$, where $1 < n < p$ and $1 < m < p$. If $a = m \cdot 1$ and $b = n \cdot 1$, then,

$$a \cdot b = (m \cdot 1)(n \cdot 1) = \left(\sum_{i=1}^{m} 1 \right) \left(\sum_{j=1}^{n} 1 \right) = mn \cdot 1 = p \cdot 1 = 0$$

This implies $a = 0$ and $b = 0$, which contradicts what we had assumed when we started. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 32 *subfield*

Let E and F be two fields, and let F be a subset of E. Then F is called a *subfield* of E if the addition and multiplication of E, when restricted to F, are the same as those of F.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 21 *elements of finite field*

A finite field F of characteristic p contains p^n elements for some integer $n \geq 1$.

Proof. Choose an element α_1 from F^*. Then

$$0 \cdot \alpha_1, \ldots, (p-1) \cdot \alpha_1$$

are pairwisely distinct from one another, for if

$$i \cdot \alpha_1 = j \cdot \alpha_1$$

for some

$$0 \leq i \leq j \leq p-1$$

then $(j-i) \cdot \alpha_1 = 0$. Since p is the characteristic of F, by Theorem 20 p can be either zero or prime. And since $0 \leq j-i \leq p-1$, therefore $j-i=0$, that is $i=j$. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Next, if

$$F \backslash \{0 \cdot \alpha_1, \ldots, (p-1) \cdot \alpha_1\}$$

is not empty we choose from it α_2. Then

$$a_1\alpha_1 + a_2\alpha_2$$

are pairwise distinct for all $0 \leq a_1, a_2 \leq p-1$, for if $a_1\alpha_1 + a_2\alpha_2$ for some $0 \leq a_1, a_2, b_1, b_2 \leq p-1$, then necessarily $a_2 = b_2$ because otherwise,

$$\alpha_2 = \frac{a_1 - b_1}{b_2 - a_2}\alpha_1$$

which contradicts the way we have chosen α_2. Then it follows that

$$(a_1, a_2) = (b_1, b_2)$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Since F is finite, we may continue in this fashion to α_3, α_4, and so on until α_n for some integer n, and find α_j, for all $2 \leq j \leq n$, from

$$F \backslash \{\sum_{i=1}^{j-1} a_i \alpha_i\}$$

where a_i, $i = 1, \ldots, j-1$, are in $\mathbf{Z}_p$.
In the end,

$$F = \{\sum_{i=1}^{n} a_i \alpha_i\}$$

where $a_1, \ldots, a_n$ are in $\mathbf{Z}_p$. In the same manner as above, we may show that

$$a_1\alpha_1 + \ldots + a_n\alpha_n$$

are pairwisely distinct from each other for all a_i in $\mathbf{Z}_p$, where $i = 1, \ldots, n$. Therefore

$$|F| = p^n$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 33 *polynomial ring*
Let F be a field. Then the set,

$$F[x] = \left\{ \sum_{i=0}^{n} a_i x^i \right\}$$

where a_i is an element in F and $n \geq 0$, is called the *polynomial ring* over F. An element of $F[x]$ is called a *polynomial* over F. For a polynomial

$$f(x) = \sum_{i=0}^{n} a_i x^i$$

providing that $a_n \neq 0$, the integer n is called the *degree* of $f(x)$, denoted by $\deg(f(x))$. We define $\deg(0) = -\infty$. A nonzero polynomial $f(x)$ of degree n is said to be *monic* if $a_n = 1$. Furthermore, a polynomial $f(x)$ is said to be *reducible* over F if there exist two polynomials $g(x)$ and $h(x)$ over F such that $\deg(g(x)) < \deg(f(x))$ and $\deg(h(x)) < \deg(f(x))$, and $f(x) = g(x)h(x)$. A polynomial is said to be *irreducible* over F if it is not reducible.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 34 *remainder of polynomial ring*

Let $f(x)$ in $F[x]$ be a polynomial of degree $n \geq 1$. Then, for any polynomial $g(x)$ in $F[x]$ there exists a unique pair $(s(x), r(x))$ of polynomials, where

$$\deg(r(x)) < \deg(f(x))$$

or $r(x) = 0$, such that

$$g(x) = s(x)f(x) + r(x)$$

Here $r(x)$ is called the *principal remainder* of $g(x)$ divided by $f(x)$, or in our notation

$$(g(x)(\mathrm{mod}\, f(x)))$$

Coding theory, Group, field and finite field, 11^{th} November 2005 –27– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 35 *the greatest common divisor and the least common multiple*

Let $f(x)$ and $g(x)$ in $F[x]$ be two nonzero polynomials. The *greatest common divisor* of $f(x)$ and $g(x)$, written $\gcd(f(x), g(x))$, is the monic polynomial of the highest degree which is a divisor of both $f(x)$ and $g(x)$. Two polynomials $f(x)$ and $g(x)$ are said to be *co-prime*, or *prime*, to each other if $\gcd(f(x), g(x)) = 1$. The *least common multiple* of $f(x)$ and $g(x)$, namely $\mathrm{lcm}(f(x), g(x))$, is the monic polynomial of the lowest degree which is a multiple of both $f(x)$ and $g(x)$.

Coding theory, Group, field and finite field, 11^{th} November 2005 –28– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 19 *factorisation*

Let the factorisations of two polynomials $f(x)$ and $g(x)$ are,

$$f(x) = a \cdot (p_1(x))^{e_1} \cdots (p_n(x))^{e_n}$$

and

$$g(x) = b \cdot (p_1(x))^{d_1} \cdots (p_n(x))^{d_n}$$

where a and b are in F^*, and $e_i, d_i \geq 0$, and $p_i(x)$ are distinct monic irreducible polynomials, then,

$$\gcd(f(x), g(x)) = (p_1(x))^{\min(e_1,d_1)} \cdots (p_n(x))^{\min(e_n,d_n)}$$

and

$$\mathrm{lcm}(f(x), g(x)) = (p_1(x))^{\max(e_1,d_1)} \cdots (p_n(x))^{\max(e_n,d_n)}$$

Coding theory, Group, field and finite field, 11^{th} November 2005 –29– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 20 *greatest common divisor*

Let $f(x)$ and $g(x)$ in $F[x]$ be two nonzero polynomials. Then, there exist two polynomials $u(x)$ and $v(x)$ having

$$\deg(u(x)) < \deg(g(x))$$

and

$$\deg(v(x)) < \deg(f(x))$$

such that,

$$\gcd(f(x), g(x)) = u(x)f(x) + v(x)g(x)$$

Then,

$$\gcd(f(x)h(x), g(x)) = \gcd(f(x), g(x))$$

if $\gcd(h(x), g(x)) = 1$.

Coding theory, Group, field and finite field, 11^{th} November 2005 –30– From 25^{th} October 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 22 *rings of polynomial*

Let $f(x)$ be a polynomial of degree n over a field F, where $n \geq 1$. Then $F[x]/(f(x))$, together with the addition,

$$g(x) \oplus h(x) = (g(x) + h(x) (\mathrm{mod}\, f(x)))$$

also written $g(x) + h(x)$, and multiplication,

$$g(x) \odot h(x) = (g(x)h(x) (\mathrm{mod}\, f(x)))$$

also written $g(x) \cdot h(x)$, form a ring. Furthermore, $F[x]/(f(x))$ is a field if and only if $f(x)$ is irreducible.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 21 *rings of polynomial*

Consider the ring $\mathbf{Z}_2[x]/(1+x^2) = \{0, 1, x, 1+x\}$ Its addition and multiplication tables are shown in Table 2.

$+$	0	1	x	$(1+x)$
0	0	1	x	$1+x$
1	1	0	$1+x$	x
x	x	$1+x$	0	1
$1+x$	$1+x$	x	1	0

$\times$	0	1	x	$1+x$
0	0	0	0	0
1	0	1	x	$1+x$
x	0	x	1	$1+x$
$1+x$	0	$1+x$	$1+x$	0

Table 2 *Addition and multiplication tables for* $\mathbf{Z}_2[x]/(1+x^2)$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 22

Consider the ring $\mathbf{Z}_2[x]/(1+x+x^2)$. Its addition and multiplication tables are given in Table 3.

$+$	0	1	x	$1+x$
0	0	1	x	$1+x$
1	1	0	$1+x$	x
x	x	$1+x$	0	1
$1+x$	$1+x$	x	1	0

$\times$	0	1	x	$1+x$
0	0	0	0	0
1	0	1	x	$1+x$
x	0	x	$1+x$	1
$1+x$	0	$1+x$	1	x

Table 3 *Addition and multiplication tables for* $\mathbf{Z}_2[x]/(1+x+x^2)$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 23 *analogies between* $\mathbf{Z}$ *and* $F[x]$

Table 4 shows the analogies between $\mathbf{Z}$ and $F[x]$.

the integral ring $\mathbf{Z}$	the polynomial ring $F[x]$
an integer m	a polynomial $f(x)$
a prime number p	an irreducible polynomial $p(x)$
$\mathbf{Z}_m = \{0, \ldots, m-1\}$	$F[x]/(f(x)) = \{\sum_{i=0}^{n-1} a_i x^i; a_i \in F, n \geq 1\}$
$a \oplus b = (a + b (\mathrm{mod}\, m))$	$g(x) \oplus h(x) = (g(x) \oplus h(x) (\mathrm{mod}\, f(x)))$
$a \odot b = (ab (\mathrm{mod}\, m))$	$g(x) \odot h(x) = (g(x)h(x) (\mathrm{mod}\, f(x)))$
$\mathbf{Z}_m$ is a ring	$F[x]/(f(x))$ is a ring
$\mathbf{Z}_m$ is a field $\Leftrightarrow$ m is a prime	$F[x]/(f(x))$ is a field $\Leftrightarrow$ $f(x)$ is irreducible

Table 4 *Analogies between* $\mathbf{Z}$ *and* $F[x]$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 23 *finite fields*

For every element ϕ of a finite field F with n elements, $\phi^n = \phi$.

Proof. The case when $\phi = 0$ is trivial. Next, if $\phi \neq 0$, then we could list all the nonzero elements of F as

$$F^* = \{\phi_1, \ldots, \phi_{n-1}\}$$

And since F is closed, we could multiply each element in F^* to obtain

$$F^* = \{\phi\phi_1, \ldots, \phi\phi_{n-1}\}$$

Therefore

$$\phi_1 \cdots \phi_{n-1} = (\phi\phi_1) \cdots (\phi\phi_{n-1})$$

which leads to

$$\phi^{n-1} = 1$$

¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Corollary 23[1] *finite fields*

Let F be a subfield of E, and let $|F| = n$. Then an element ϕ of E is also in F if and only if $\phi^n = \phi$.

Proof. The *if* part was already proved in Theorem 23. For the *only if* part, if ϕ satisfy $\phi^n = \phi$, then it is a root of $x^n - x$. And since $|F| = n$ means that all the elements of F are roots of $x^n - x$, it follows that ϕ lies in F. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 36 *finite field*

We denote a finite field with q elements by $\mathbf{F}_q$ or $GF(q)$. Let α be a root of an irreducible polynomial $f(x)$ of degree n over a field F. Then, if we replace x in $F[x]/(f(x))$ by α, the field $F[x]/(f(x))$ can be represented as,

$$F[\alpha] = \left\{ \sum_{i=0}^{n-1} a_i \alpha^i \right\}$$

for a_i in F.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 37 *primitive element*

An element α in a finite field $\mathbf{F}_q$ is called a *primitive element*, or *generator*, of $\mathbf{F}_q$ if

$$\mathbf{F}_q = \{0, \alpha, \alpha^2, \ldots, \alpha^{q-1}\}$$

§

Definition 38 *order*

The *order*, $\mathrm{ord}(\alpha)$, of a nonzero element α in $\mathbf{F}_q$ is the smallest positive integer k such that $\alpha^k = 1$.

§

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 39 *prime power*

A *prime power* is a prime or an integer power of a prime.

§

Example 24 *prime power*

Examples of prime powers are,

$$2, 3, 4, 5, 7, 8, 9, 11, 13, 16, 17, \ldots$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 40 *linear code*

Let the alphabet be $\mathbf{F}_q$, in other words a Galois field $GF(q)$, where q is a prime power, and let the vector space $V(n, q)$ be $(\mathbf{F}_q)^n$. Then a *linear code* over $GF(q)$, for some positive integer n, is a subspace of $V(n, q)$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 24 *linear code*

A subset C of $V(n,q)$ is a linear code if and only if,

a. $\mathbf{u}+\mathbf{v} \in C$ for all $\mathbf{u}$ and $\mathbf{v}$ in C

b. $a\mathbf{u} \in C$ for all $\mathbf{u} \in C$ and $a \in GF(q)$

Proof. The proof follows from Definition 40 since, if C is a field, it must be closed under addition and multiplication. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 25 *linear binary code*

A binary code is linear if and only if the sum of any two code words is a code word.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 41 *vector space*

A *vector space* V is a set which is closed under finite vector addition and scalar multiplication. If the scalars are members of a field F, then V is called a vector space under F. Furthermore, V is a vector space under F if and only if for all members of V and F the following properties hold under addition,

a. commutativity
b. associativity
c. existence of an identity
d. existence of an inverse

while under multiplication the following,

e. associativity under scalar multiplication
f. distributivity of scalar sum
g. distributivity of vector sum
h. existence of a scalar multiplication identity

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

In other words, for all $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ in V and all p and q in F,

a. $\mathbf{x}+\mathbf{y}=\mathbf{y}+\mathbf{x}$
b. $(\mathbf{x}+\mathbf{y})+\mathbf{z}=\mathbf{x}+(\mathbf{y}+\mathbf{z})$
c. $\mathbf{0}+\mathbf{x}=\mathbf{x}+\mathbf{0}=\mathbf{x}$
d. $\mathbf{x}+(-\mathbf{x})=\mathbf{0}$
e. $r(s\mathbf{x})=(rs)\mathbf{x}$
f. $(r+s)\mathbf{x}=r\mathbf{x}+s\mathbf{x}$
g. $r(\mathbf{x}+\mathbf{y})=r\mathbf{x}+r\mathbf{y}$
h. $1\mathbf{x}=\mathbf{x}$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 26 *vector space over finite field*

Let q be a prime power, and let $GF(q)$ denote a finite field over q elements. Then, by *vector space over finite field* we mean a set $GF(q)^n$ of all ordered n-tuples over $GF(q)$, which is closed under finite vector addition and multiplication, that is to say, multiplication by some scalar a in $GF(q)$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 25 *vector space*

A non-empty subset C of $V(n,q)$ is a subspace if and only if C is closed under addition and scalar multiplication. In other words

Proof. What Theorem 25 states amounts to saying that a non-empty C in $V(n,q)$ is a subspace if and only if,

a. $\mathbf{x}, \mathbf{y} \in C$ implies $\mathbf{x} + \mathbf{y} \in C$

b. if $a \in GF(q)$ and $\mathbf{x} \in C$, then $a\mathbf{x} \in C$

All properties to be met in Definition 41 are the same for C as for $V(n,q)$ itself, provided that C is closed under addition and scalar multiplication. Therefore statements (a) and (b) are necessary for C to be a subspace. They are also sufficient since C is already a subset of $V(n,q)$. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 42 *linearly independence*

A *linear combination* of r vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$ in $V(n,q)$ is any vector of the form $\sum_{i=1}^{r} a_i \mathbf{v}_i$, where a_i are scalars. Let A be a set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$. Then A is said to be *linearly dependent* if there exist scalars $a_1, \ldots, a_r$ not all of which are zero, such that

$$\sum_{i=1}^{r} a_i \mathbf{v}_i = \mathbf{0}$$

And A is *linearly independent* if it is not linearly dependent, that is to say, if $\sum_{i=1}^{r} a_i \mathbf{v}_i = \mathbf{0}$ implies a_i are all zero for $i = 1, \ldots, r$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 43 *generating set and basis*

Let C be a subspace of a vector space $V(n,q)$ over $GF(q)$. Then a subset $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ of C is called a *generating-* or *spanning set* of C if every vector in C can be expressed as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_r$.

A *basis* of C is a generating set of the same which is also linearly independent.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 44 *relative minimum distance*

For a q-ary (n, m, d)-code C, the *relative minimum distance* of C is defined to be

$$\delta(C) = \frac{d-1}{n}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 45 *optimal code*

Let a code alphabet A be of size $q > 1$, n the size of each word, d the minimum distance, and $A_q(n, d)$ the largest possible vocabulary size m such that there exists an (n, m, d)-code over A. Then any (n, m, d)-code C which has $m = A_q(n, d)$ is called an *optimal code*.

The *main coding theory problem* is precisely to find the value of $A_q(n, d)$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 46 *Hamming sphere*

Consider each word as an n-tuple. Then all such tuples lying within Hamming distance r of an n-tuple x are said to be within a *Hamming sphere* of radius r around x.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 26 *Hamming bound*

Let the size of the alphabet be $q = |A|$, the size of a word be n, and the Hamming- or minimum distance be d. Then the Hamming- or sphere-packing bound on the size m of a code dictionary C is given by,

$$m \leq \frac{q^n}{\sum_{i=0}^{r^90} (q-1)^i \binom{n}{i}}$$

where

$$r^90 = \left\lfloor \frac{d-1}{2} \right\rfloor$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proof. Let c be a code word. Let $e(x, y)$ be the number of places which are different between two words x and y. Since there are $q-1$ possibilities for each differing position between any two words, there are $(q-1)^i$ possible errors when i places are different. And to position these i places there are altogether $\binom{n}{i}$ ways. Therefore the number of all words w_i such that $e(w_i, c) \leq r$ is the number n_r of n-tuples in a Hamming sphere of radius r around c, and is,

$$n_r = \sum_{i=0}^{r} (q-1)^i \binom{n}{i}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Then the lower bound for our code is $d(C) > 2r$, that is to say, $d(C) \geq 2r+1$. In other words, Hamming spheres of radius r around the m code words of C are mutually nonintersecting. There are a total of q^n possible n-tuples, that is words of length n, not all of which are code words. In other words, $m < q^n$. And since there are n_r of these n-tuples within each sphere, the the number of the all the n-tuples contained within the space of all these n-tuples over the alphabet A is $n_r m$. Hence,

$$m \sum_{i=0}^{r} (q-1)^i \binom{n}{i} \leq q^n$$

and thus this theorem. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 47 *perfect code*
Codes which satisfies the Hamming bound are called *perfect codes.*

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Problem 1 *Stirling's approximation of* $n!$
Let r and n be integers such that $0 < r < \frac{n}{2}$, then prove that,

$$\left[8n\left(\frac{r}{n}\right)\left(1-\frac{r}{n}\right)\right]^{-\frac{1}{2}} 2^{nH\left(\frac{r}{n},1-\frac{r}{n}\right)} \leq \sum_{i=0}^{r}\binom{n}{i} \leq 2^{nH\left(\frac{r}{n},1-\frac{r}{n}\right)}$$

where $H(x,y)$ is the entropy function the arguments x and y of which are probabilities and $H(\cdot,\cdot)$ has the unit of bits per symbol. (Hint: Stirling's approximation to $n!$, *cf* MacWilliams and Sloane, 1977)

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Note 1 *dictionary's size*

Let $C(n,d)$ be a code with words of length n and minimum distance between words d. Let $m_{n,d}$ be the number of code words in $C(n,d)$. Then the size of the largest dictionary of n-tuples with fractional minimum distance d_f is,

$$m_m(n,d_f) = \max_{\{C(n,d):\left(\frac{d}{n}\right)\geq d_f\}} |C(n,d)|$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Problem 2

From Note 1, show that for n fixed, $m_m(n,d_f)$ is a monotonous nonincreasing function of d_f. Then show that with d_f fixed, $m_m(n,d_f)$ increases exponentially with n.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 48 *asymptotic transmission rate*

The *asymptotic transmission rate* is defined to be,

$$R(d_f) = \lim_{n\to\infty} \frac{1}{n} \log m_m(n, d_f)$$

Also defined are the upper- and the lower bounds on this rate,

$$\bar{R}(d_f) = \limsup_{n\to\infty} \frac{1}{n} \log m_m(n, d_f)$$

and

$$\underline{R}(d_f) = \liminf_{n\to\infty} \frac{1}{n} \log m_m(n, d_f)$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Note 2 *bounds on rate*

For large n, show that $\underline{R}\,(d_f) < R\,(d_f) < \bar{R}\,(d_f)$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 27 *Hamming bound for binary codes*

Using the results from Problem 1 we obtain the Hamming bound for the binary code,

$$m \leq \left(8n\left(\frac{r}{n}\right)\left(1-\frac{r}{n}\right)\right)^{\frac{1}{2}} 2^{n\left(1-H\left(\frac{r}{n},1-\frac{r}{n}\right)\right)} \tag{16}$$

where

$$r = \left\lfloor \frac{d-1}{2} \right\rfloor$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Equation 16 must hold for all binary dictionaries, therefore it gives an upper bound on the maximum dictionary size $m_m(n, d_f)$ over all dictionaries whose word length is n and fractional distance,

$$d_f = \frac{d}{n} = \frac{2r + \left\{ \begin{matrix} 1 \\ 2 \end{matrix} \right\}}{n}$$

where the choice of 1 or 2 depends on whether d is odd or respectively even. For large n,

$$m_m(n, d_f) \leq \left(9n\left(\frac{d_f}{2}\right)\left(1-\frac{d_f}{2}\right)\right)^{\frac{1}{2}} 2^{n\left(1-H\left(\frac{d_f}{2},1-\frac{d_f}{2}\right)\right)}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

The upper bound for the attainable information rate is,

$$\begin{aligned}\bar{R}(d_f) &= \limsup_{n\to\infty} \frac{1}{n}\log_2 m_m(n,d_f) \\ &\leq \lim_{n\to\infty}\left\{\frac{1}{2}\frac{\log_2 n}{n} + \frac{1}{2n}\log_2\left(\frac{9d_f}{2}\left(1-\frac{d_f}{2}\right)\right)\right\} + 1 - H\left(\frac{d_f}{2}, 1-\frac{d_f}{2}\right)\end{aligned}$$

As n approaches infinity,

$$\bar{R}(d_f) \leq 1 - H\left(\frac{d_f}{2}, 1-\frac{d_f}{2}\right)$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 27 *Plotkin's bound*

Let $d(c_i, c_j)$ be the Hamming distance between the code words c_i and c_j. Let $d(C)$ be the minimum distance between code words, and $\bar{d}$ the average distance between words. If,

$$\frac{d}{n} > \frac{q-1}{q}$$

then the *Plotkin's bound*,

$$m_{n,d} \leq \frac{\frac{d}{n}}{\frac{d}{n} - \frac{q-1}{q}}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proof. The average distance gives an upper bound for the minimum distance, that is $d \leq \bar{d}$, where

$$\begin{aligned}\bar{d} &= \frac{\sum_{i=2}^{m}\sum_{j=1}^{i-1} d\left(c_i, c_j\right)}{\sum_{i=2}^{m}\sum_{j=1}^{i-1} 1} \\ &= \left(\frac{m(m-1)}{2}\right)^{-1} \sum_{i=2}^{m}\sum_{j=1}^{i-1} d\left(c_i, c_j\right)\end{aligned}$$

Since the Plotkin's bound is an upper bound on d, we need to maximise,

$$\sum_{i>j}\sum d\left(c_i, c_j\right) = \sum_{i>j}\sum\sum_{k=1}^{n} d\left(c_{ik}, c_{jk}\right) = \sum_{k=1}^{n}\sum_{i>j}\sum d\left(c_{ik}, c_{jk}\right)$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

This implies (*cf* Plotkin, 1960),

$$\sum_{i>j}\sum d\left(c_i, c_j\right) \leq \sum_{k=1}^{n} \max_{\{c_{ik}, i=1,\ldots,m\}} \left\{\sum_{i>j}\sum d\left(c_{ik}, c_{jk}\right)\right\}$$

which says that the upper bound is maximised by choosing a maximising c_{ik} from the alphabet A. However this is,

$$\max_{c_{ik}, i=1,\ldots,m} \sum_{i>j}\sum d\left(c_{ik}, c_{jk}\right) \leq \left(\frac{m}{q}\right)^2 \frac{q(q-1)}{2}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Providing that,

$$\frac{d}{n} > \frac{q-1}{q}$$

then

$$d \leq n\left(\frac{m}{m-1}\right)\left(\frac{q-1}{q}\right)$$

¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Note 3 *double summations*

Notice how,

$$\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}(\cdot) = \sum_{i=2}^{m}\sum_{j=1}^{i-1}(\cdot)$$

Equivalently to this are,

$$\sum\sum_{i<j}(\cdot) \quad \text{and} \quad \sum\sum_{i>j}(\cdot)$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Note 4 *Plotkin's bound*

If,

$$d_f > \frac{q-1}{q}$$

then,

$$m_m(n, d_f) \leq \frac{d_f}{d_f - \left(\frac{q-1}{q}\right)}$$

and then,

$$\bar{R}(d_f) = \limsup_{n\to\infty} \frac{1}{n} \log m_m(n, d_f) = 0$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

On the other hand if,

$$d_f \leq \frac{q-1}{q}$$

then from,

$$m(n, d) = \sum_{a\in A} m_a(n, d)$$

where $m(n, d) = |C(n, d)|$, $C(n, d)$ being any code consisting of n-tuples whose minimum distance is at least d, and $m_x(n, d) = |C_x(n, d)|$, $C_x(n, d)$ comprising all n-tuples in $C(n, d)$ which begin with the symbol x. Hence,

$$\begin{aligned} m(n, d) \leq q m_x(n, d) &= q m(n-1, d) \\ &\vdots \\ &= q^{n-k} m(k, d) \end{aligned}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Provided k is small enough, we may yet use the Plotkin's bound, hence

$$m(n,d) \leq \frac{q^{n-k}\left(\frac{d}{k}\right)}{\left(\frac{d}{k}\right) - \left(\frac{q-1}{q}\right)}$$

when

$$\frac{d}{k} > \frac{q-1}{q}$$

Choose k the largest integer satisfying

$$\frac{d}{k} - \frac{1}{qk} \geq \frac{q-1}{q}$$

Coding theory, Bounds in coding, 18^{th} November 2005 –33– From 15^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Then,

$$k + r = \frac{qd-1}{q-1}$$

where $0 \leq r < 1$. And then,

$$m(n,d) \leq \frac{q^{n-\left(\frac{qd-1}{q-1}\right)r+1}d}{(q-1)r+1}$$

Finally,

$$m(n,d) \leq q^{n-\left(\frac{qd-1}{q-1}\right)}d$$

and, if d_f is fixed and n become large,

$$\bar{R}(d_f) \leq \log q \left(1 - \frac{q}{q-1}d_f\right)$$

Coding theory, Bounds in coding, 18^{th} November 2005 –34– From 15^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Problem 5 *upper bound*

Prove that,

$$q^r \left((q-1)\, r + 1 \right)^{-1} \leq 1$$

for $0 \leq r \leq 1$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proposition 1 *Elias bound for binary alphabet*

Let C be a code containing binary n-tuples, $m_d(x)$ the number of code words within distance d of an n-tuple x. Further, let A be a new code whose code words are the difference vectors $a_1, \ldots, a_{m_d}$ such that $a_i = c_i \ominus x$, $i = 1, \ldots, m_d$, where $\ominus$ denotes modulo subtraction of the vectors, element by element. Assume that $d < \frac{n}{2}$ and both d and m are large enough such that $m_d(x) \geq 2$. Then,

$$\frac{d_c}{n} \leq \frac{2d}{n} \left(1 - \frac{d}{n} \right) \frac{m_a}{m_a - 1} \tag{17}$$

where

$$m_a \geq \left\lceil 2^{-n} m \sum_{i=0}^{d} \binom{n}{i} \right\rceil$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proof. Since C is a code of binary n-tuples, there are

$$\sum_{i=0}^{d} \binom{n}{i}$$

n-tuples within distance d of each code word. This gives the total of

$$m \sum_{i=0}^{d} \binom{n}{i}$$

n-tuples in the Hamming sphere around the m code words.
There are $m_d(x)$ code words within the distance d of any n-tuple x. For x in X^n, c in C and $d(x,c) \leq d$, the number of pairs (x,c) can be counted by picking up first x and then c, hence

$$\sum_{x \in X^n} m_d(x) = m \sum_{i=0}^{d} \binom{n}{i}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Since X^n contains 2^n of n-tuples, consequently there exists some value of x such that,

$$m_d(x) \geq \left\lceil 2^{-n} m \sum_{i=0}^{d} \binom{n}{i} \right\rceil$$

Let $c_1, \ldots, c_{m_d}$ be code words in C that lie within Hamming distance d of the n-tuple x. Consider the difference vector $a_1, \ldots, a_{m_d}$ such that $a_i = c_i \ominus x$. Then A is a set of localised code words of C. Then,

$$a_i \ominus a_j = (c_i \ominus x) \ominus (c_j \ominus x) = c_i \ominus c_j$$

and we have,

$$d(c_i, c_j) = d(a_i, a_j)$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Thus,

$$m_a \geq m_d(x) \geq \left\lceil 2^{-n} m \sum_{i=0}^{d} \binom{n}{i} \right\rceil$$

Also, $d_a \geq d_c$ and $w(a_i) \leq d$ for all n-tuple a_i in A, where the Hamming weight $w(a_i)$ is the number of nonzero elements in a_i.
Next, applying the average-distance Plotkin bound to the localised code A one obtains,

$$d_c \leq d_a \leq \bar{d}_a = \left(\frac{m_a(m_a-1)}{2} \right)^{-1} \sum_{i>j} \sum d(a_i, a_j) \tag{18}$$

We maximise RHS of Equation 18 to get rid of the dependence on A. We enlarge our restriction on $w(a_i)$ above to the set of all possible a_i in A, thus,

$$\sum_{a_i \in A} w(a_i) \leq m_a d \tag{19}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Then, let z_k be the number of code words in A having a 0 in the k^{th} position. We maximise,

$$\sum_{i>j} \sum d(a_i, a_j) = \sum_{k=1}^{n} (m_a - z^k) \tag{20}$$

subject to the constraint of Equation 19 that,

$$\sum_{k=1}^{n} (m_a - z_k) \leq m_a d \tag{21}$$

By setting,

$$z_k = \frac{m_a d}{n} \tag{22}$$

we maximise RHS of Equation 20 under the constraint in Equation 21. From Equation's 18, 20 and 22 we obtain Equation 17. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Algorithm 2 *Gilbert bound, a lower bound to m for n, d and q.*

$S^n \leftarrow X^n$
for all c_i in S^n **do**
  for all n-tuples c_j within $d-1$ distance of C **do**
    remove c_j
  endfor
endfor

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Note 5 *Gilbert bound*

For the Gilbert bound algorithm, Algorithm 2, initially $|S|^n = |X|^n$. For each c_i chosen, at most

$$\sum_{i=0}^{d-1}(q-1)^i \binom{n}{i}$$

n-tuples are removed. If

$$(m-1)\sum_{i=0}^{d-1}(q-1)^i \binom{n}{i} < q^n$$

then the algorithm will not stop after $m-1$ code-word selections.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 49 *Group, ring and field*

A non-empty set G with a binary composition is called a *group* if the composition is associative, if a unique *identity* exists for all elements in G, and if a unique *inverse* exists for each of the elements in G. The group G is called *Abelian* if the composition in it is commutative for any two elements in G. A non-empty set R with two binary compositions, call these addition and multiplication, defined on it is called a *ring* if R is an Abelian group with respect to the composition addition, if multiplication in R is associative, and if distributive laws hold for all elements in R. A set F having at least two elements with two compositions, be them called addition and multiplication, defined on it is called a *field* if it is a commutative ring with identity every non-zero element of which has an inverse with respect to multiplication. A field having only a finite number of elements is called a *finite* or *Galois field.*

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –1–
From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 28 *finite field*

The set

$$F_p = \{0, \ldots, p-1\}$$

in which addition and multiplication are defined modulo p, where p is a prime integer, is a finite field. For $p = 2$ we have $F_2 = \{0, 1\}$, which is denoted by $\mathbf{B}$. The set $\mathbf{B}^n$ of all ordered n-tuples or sequences of length n, a positive integer, with each tuple or entry of the sequence being in the field $\mathbf{B}$ and a composition defined as a componentwise summation of any two sequences in $\mathbf{B}^n$, is an Abelian group. The zero sequence of length n is the identity of $\mathbf{B}^n$ and each element in $\mathbf{B}^n$ is its own inverse.

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –2–
From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 50 *code words*

A *binary block* (b, n)*-code* comprises an *encoding function*

$$E : \mathbf{B}^b \to \mathbf{B}^n$$

and a *decoding function*

$$D : \mathbf{B}^n \to \mathbf{B}^b$$

The images of E are called *code words.*

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –3–
From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 51 *distance function*

Let two binary sequences be a and b in $\mathbf{B}^n$. The *distance* $d(a, b)$ between a and b is defined as

$$d(a, b) = \sum_{i=1}^{n} x_i$$

where

$$x_i = \begin{cases} 0 & \text{if } a_i = b_i \\ 1 & \text{if } a_i \neq b_i \end{cases}$$

Definition 52 *weight function*

The *weight* $w(a)$ of a in $\mathbf{B}^n$ is the number of non-zero components of the sequence a.

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –4–
From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 28 *weight and distance*

Let a and b be any two sequences in $\mathbf{B}^n$. Then $d(a,b) = w(a+b)$.

Proof. The only contribution of 1 to $d(a,b)$ is $a_i \neq b_i$ for all $1 \leq i \leq n$. But this latter is the case if and only if $a_i + b_i = 1$, and this contributes 1 to $w(a+b)$. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 53 *homomorphism*

Let X and Y be two groups. Then a map

$$f : X \to Y$$

which satisfies the property

$$f(x_1 x_2) = f(x_1) f(x_2)$$

for all x_1 and x_2 in X is called a *homomorphism*. Further, the homomorphism f is called a *monomorphism* if it is one to one, and it is called an *isomorphism* if it is both one to one and onto.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 54 *group code*

A block code is called a *group code* if all its code words form an additive group.

Definition 55 *generator matrix*

A $b \times n$ matrix G over $\mathbf{B}$, where $b < n$, is called an *encoding-* or *generator matrix* if G is of the form

$$G = [I_b \, G_n]$$

where I_b is an identity matrix of dimension b and G_n a $b \times (n - b)$ matrix. An *encoding function* $E : \mathbf{B}^b \to \mathbf{B}^n$ is defined by

$$E(x) = xG$$

for all x in $\mathbf{B}^b$

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –7–
From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 29 *monomorphic generator matrix*

The encoding function $E : \mathbf{B}^b \to \mathbf{B}^n$ given by $E(x) = xG$ for all x in $\mathbf{B}^b$, where G is a $b \times n$ generator matrix, is a monomorphism.

Proof. Both $\mathbf{B}^b$ and $\mathbf{B}^n$ are additive Abelian groups. Then for all x and y in $\mathbf{B}^b$ we know that $x + y$ is also in $\mathbf{B}^b$ and

$$E(x + y) = (x + y)G = xG + yG = E(x) + E(y)$$

Thus E is a homomorphism. Further, as the first part of G is I_b, it follows that a part of $E(x)$ is x itself. Therefore the matrix encoding method gives for each binary message word a distinct code word. In other words, the mapping E is one to one, which means that it is a monomorphism. ¶

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –8–
From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 56 *matrix code*

A code generated by a generating matrix is called a *matrix code.*

Theorem 30 *matrix code a group code*

A matrix code is a group code.

Proof. The code words generated by E are associative, since

$$x_1G + (x_2G + x_3G) = (x_1G + x_2G) + x_3G$$

They have a unique identity, that is the zero $b \times n$ matrix, and each of them is its own inverse. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 57 *parity check code*

An $(b, b+1)$ parity check code is the code generated by an encoding function $E : \mathbf{B}^b \to \mathbf{B}^{b+1}$ defined by

$$E(a_1 \cdots a_b) = a_1 \cdots a_b a_{b+1}$$

where

$$a_{b+1} = \begin{cases} 1 & \text{if } w(a) \text{ is odd} \\ 0 & \text{if } w(a) \text{ is even} \end{cases}$$

$w(a)$ being $w(a_1 \cdots a_b)$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 31 *parity check code*

The $(b, b+1)$ parity check code is a group code.

Proof. Let our unencoded binary words be $a = a_1 \cdots a_b$, $b = b_1 \cdots b_b$, and $c = c_1 \cdots c_b$ such that $c_i = a_i + b_i$ for $i = 1, \ldots, b$, and let the coded words of a and b be respectively $\bar{a} = aa_{b+1}$ and $\bar{b} = bb_{b+1}$. Since c is odd if and only if either a is odd while b is even or vice versa, but when this is the case we have either $a_{b+1} = 1$ and $b_{b+1} = 0$, or $a_{b+1} = 0$ and $b_{b+1} = 1$. Either way we have

$$c_{b+1} = 1 = a_{b+1} + b_{b+1}$$

Next, c is even if and only if a and b are either both odd or both even. But when either of these is the case, then

$$a_{b+1} + b_{b+1} = 0 = c_{b+1}$$

Hence $\bar{c}$ is a parity-check code word. The zero word is the identity and the inverse of each word is that word itself. Therefore the set of all code words forms a group. ¶

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –11– From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 32 *minimum distance*

The minimum distance of a group code equals the minimum of the weights of its non-zero code words.

Proof. Let d_m be the minimum distance of the group code, and w_m the minimum of the weights of the non-zero code words of the same. Then there exist code words a and b such that

$$d_m = d(a, b) = w(a + b) \geq w_m$$

Now, w_m implies that there exists a non-zero code word c such that

$$w_m = w(c) = d(c, 0) \geq d_m$$

Hence $d_m = w_m$. ¶

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –12– From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 29 *generator and parity check matrices*

Let the generator matrix be

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

The dimension of G is $b \times n$, which in this case is 3×6. Let $a_1a_2a_3a_4a_5a_6$ be the code word and $a_1a_2a_3$ the original word, then

$$(a_1\, a_2\, a_3\, a_4\, a_5\, a_6) = (a_1\, a_2\, a_3)\, G$$

and then,

$$a_4 = a_1 + a_2$$
$$a_5 = a_1 + a_3$$
$$a_6 = a_1 + a_2 + a_3$$

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –13– From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

In other words,

$$\left.\begin{aligned} a_1 + a_2 + a_4 &= 0 \\ a_1 + a_3 + a_5 &= 0 \\ a_1 + a_2 + a_3 + a_6 &= 0 \end{aligned}\right\} \text{parity check equations}$$

These parity check equations are then, in matrix form,

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{pmatrix} = 0$$

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –14– From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

The matrix

$$H = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

is called the *parity check matrix* of the code. Then $G = (\, I_3 \quad A \,)$ and $H = (\, A^9 0 \quad I_3 \,)$, where

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

and

$$A^9 0 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –15– From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 30

The parity check code in Definition 57 is in fact a matrix code given by the generator matrix

$$G = \begin{pmatrix} 1 & 0 & \cdots & 0 & 1 \\ 0 & 1 & & 0 & 1 \\ \vdots & & \ddots & & \vdots \\ 0 & & \cdots & 1 & 1 \end{pmatrix}$$

whose parity check matrix is the $1 \times (b+1)$ matrix $H = (\, 1 \quad \cdots \quad 1 \,)$.

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –16– From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 58 *syndrome*
The *syndrome* of a word $r \in \mathbf{B}^n$ is

$$\mathbf{s} = H\mathbf{r}^9 0$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Algorithm 3 *the syndrome decoding algorithm.*

$r \leftarrow r_1 \cdots r_b r_{b+1} \cdots r_n$
$\mathbf{s} \leftarrow H\mathbf{r}^9 0$
if $\mathbf{s} = 0$ **then**
  $b_r \leftarrow (r_1 \cdots r_b)$
elseif $\mathbf{s}$ matches the i^{th} column of H **then**
  $c_r \leftarrow (r_1 \cdots r_{i-1}(r_i + 1)r_{i+1} \cdots r_n)$
  $b_r \leftarrow (c_{r1} \cdots c_{rb})$
else
  at least two errors have occurred in the transmission
endif

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 33 *parity check*

An $(n-b)\times b$ parity check matrix H will decode all single errors correctly if and only if the columns of H are distinct and non-zero.

Proof. Suppose the i^{th} column of H is zero, and let e be a word whose weight is 1 having 1 in the i^{th} position and 0 elsewhere. Then for any code word b, we have

$$H(\mathbf{b}+\mathbf{e})^9 0 = H\mathbf{b}^9 0 + H\mathbf{e}^9 0 = 0$$

So our decoding procedure becomes $D(b+e)=b+e$ and the error vector $\mathbf{e}$ goes undetected.

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –19–
From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Next, suppose that the i^{th} and the j^{th} columns of H are identical. Let e^i and e^j be words of length n with 1 in the i^{th} and respectively j^{th} position and 0 elsewhere. Then for any code word b, we have $H(\mathbf{b}+\mathbf{e}^i)^9 0 =$

$$H\mathbf{b}^9 0 + H\left(\mathbf{e}^i\right)^9 0 = H\left(\mathbf{e}^i\right)^9 0 = H\mathbf{b}^9 0 + H\left(\mathbf{e}^j\right)^9 0 = H\left(\mathbf{b}+\mathbf{e}^j\right)^9 0$$

We are unable to decide whether the error occurred in the i^{th} or the j^{th} position.
Conversely, suppose all the columns of H are distinct and non-zero. Then for any code word b and any error vector $\mathbf{e}$ of weight 1 having 1 in the i^{th} position,

$$H\left(\mathbf{b}+\mathbf{e}\right)^9 0 = H\left(\mathbf{b}^9 0 + \mathbf{e}^9 0\right) = H\mathbf{b}^9 0 + H\mathbf{e}^9 0 = 0 + H\mathbf{e}^9 0$$

Our decoding procedure gives $D(b+e)=b$, therefore every single error is corrected. ¶

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –20–
From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 34 *generator and parity-check matrices*

If

$$G = (\, I_b \quad A \,)$$

is a $b \times n$ generator matrix of a code, then

$$H = (\, A^t 0 \quad I_{n-b} \,)$$

is the unique parity check matrix for the same code. If

$$H = (\, B \quad I_{n-b} \,)$$

is an $(n-b) \times n$ parity check matrix, then

$$G = (\, I_m \quad B^t 0 \,)$$

is the unique generator matrix for the same code.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proof. Let the original word be $a \in \mathbf{B}^b$ and c be the code word corresponding to a with respect to the code given by the generator matrix G. Then $\mathbf{c} = \mathbf{a}G$ Let a be $a_1 \cdots a_b$. Since the first b columns of G is an identity matrix, it follows from $\mathbf{c} = \mathbf{a}G$ that $a_i = b_i$ for all $1 \leq i \leq b$. Let $\bar{c} = c_{b+1} \cdots c_n$, then $c = c_1 \cdots c_b c_{b+1} \cdots c_n$ and $\mathbf{c} = (\, \mathbf{a} \quad \bar{\mathbf{c}} \,)$. Then,

$$\begin{aligned} H\mathbf{c}^t 0 &= (\, A^t 0 \quad I_{n-b} \,)\, (\, \mathbf{a}G \,)^t\, 0 = (\, A^t 0 \quad I_{n-b} \,)\, G^t 0 \mathbf{a}^t 0 \\ &= (\, A^t 0 \quad I_{n-b} \,)\, (\, I_m A \,)^t\, 0 \mathbf{a}^t 0 = (\, A^t 0 \quad I_{n-b} \,) \begin{pmatrix} I_m \\ A^t 0 \end{pmatrix} \mathbf{a}^t 0 \\ &= (\, A^t 0 I_m + I_{n-b} A^t 0 \,)\, \mathbf{a}^t 0 = (\, A^t 0 + A^t 0 \,)\, \mathbf{a}^t 0 \\ &= 0 \times \mathbf{a}^t 0 = 0 \end{aligned}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Therefore c is the code word corresponding to the original word a in the code given by the parity check matrix H.

Now, suppose first that c is the code word corresponding to the original word a as above in the code obtained from the parity check matrix $H = (\, A^{\mathcal{G}}0 \quad I_{n-b}\,)$. Then $c_i = a_i$ for all $1 \leq i \leq b$ and $H\mathbf{c}^{\mathcal{G}}0 = 0$. Let $\bar{c} = c_{b+1} \cdots c_n$. Then,

$$H \begin{pmatrix} \mathbf{a} \\ \bar{\mathbf{c}}^{\mathcal{G}}0 \end{pmatrix} = 0$$

$$(\, A^{\mathcal{G}}0 \quad I_{n-b}\,) \begin{pmatrix} \mathbf{a} \\ \bar{\mathbf{c}}^{\mathcal{G}}0 \end{pmatrix} = 0$$

$A^{\mathcal{G}}0\mathbf{a}^{\mathcal{G}}0 + I_{n-b}\bar{\mathbf{c}}^{\mathcal{G}}0 = 0$. Therefore $\bar{c} = \mathbf{a}A$, and

$$\mathbf{c} = (\, \mathbf{a} \quad \bar{c}\,) = (\, \mathbf{a}I_m \quad \mathbf{a}A\,) = \mathbf{a}\,(\, I_m \quad A\,) = \mathbf{a}G$$

Hence c is the code word corresponding to the original word a in the code defined by the generator matrix G. So far we have proved that codes determined by G and H are identical.

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –23– From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Suppose that to $G = (\, I_m \quad A\,)$ corresponds another parity check matrix $H_1 = (\, B \quad I_{n-b}\,)$. Let e^i be the original word with 1 in the i^{th} position and 0 elsewhere. The corresponding code word is $\mathbf{e}^i G$, that is the i^{th} row of G, or we may write $\mathbf{e}^i G = (\, \mathbf{e}^i \quad \tilde{\mathbf{e}}^i\,)$, where $\tilde{\mathbf{e}}^i$ is the i^{th} row of A. Since H_1 is a parity check matrix of the code defined by G, it follows that, $H_1\,(\, \mathbf{e}^i \quad \tilde{\mathbf{e}}^i\,)^{\mathcal{G}}0 = 0$,

$$(\, B \quad I_{n-b}\,) \begin{pmatrix} \left(\mathbf{e}^i\right)^{\mathcal{G}}0 \\ \left(\tilde{\mathbf{e}}^i\right)^{\mathcal{G}}0 \end{pmatrix} = 0$$

$$B\left(\mathbf{e}^i\right)^{\mathcal{G}}0 + \left(\tilde{\mathbf{e}}^i\right)^{\mathcal{G}}0 = 0$$

Therefore $\left(\tilde{\mathbf{e}}^i\right)^{\mathcal{G}}0$ matches the i^{th} column of B, or equivalently $\tilde{\mathbf{e}}^i$ matches the i^{th} row of $B^{\mathcal{G}}0$. Then the i^{th} row of A is identical to the i^{th} column of B. And this is true for all $1 \leq i \leq b$, so we have $B = A^{\mathcal{G}}0$ and therefore $H_1 = H$. Hence, to a given G there corresponds a unique $H = (\, A^{\mathcal{G}}0 \quad I_{n-b}\,)$. Similar argument also holds if we start with a parity check matrix H given. ¶

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –24– From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 59 *dual codes*

Let C be a (b, n) code obtained from the generator matrix

$$G = [I_b \, A]$$

Then an $(n - b, n)$ matrix code defined by the parity check matrix

$$H = [A \, I_b]$$

is called the *dual code* $C^{\perp}$ of C.

Definition 60coset

Two words x and y are said to be in the same coset if and only if $y = x + c$ for some code word c in C.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 35 *cosets*

Two words x and y in $\mathbf{B}^n$ are in the same coset of C if and only if they have the same syndrome.

Proof. By Definition 60 x and y are in the same coset if and ony if

$$y = x + c$$

for some c in C, which in turn is true if and only if $x + y = c$ in C. Then it follows from this that, $H(\mathbf{x} + \mathbf{y})^9 0 = 0$

$$H\left(\mathbf{x}^9 0 + \mathbf{y}^9 0\right) = 0$$
$$H\mathbf{x}^9 0 + H\mathbf{y}^9 0 = 0$$
$$H\mathbf{x}^9 0 = H\mathbf{y}^9 0$$

¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 61 *vector space*

Let F be a field. Then a non-empty set V is called a *vector space* over F if V and an addition form an Abelian group; for every a in F and v in V there is a uniquely defined element av in V such that for any v, v_1 and v_2 in V and any a and b in F,

$$a\left(v_1+v_2\right)=av_1+av_2$$

$$(a+b)v=av+bv$$

$$(ab)v=a(bv)$$

and

$$1v=v$$

1 being the identity of F.

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –27–
From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 62 *linear-dependence*

Let V be a vector space over a field F. Then a set $\{v_1,\ldots,v_n\}$ of elements v_i in V is said to be *linearly independent* if

$$a_1v_1+\cdots+a_nv_n=0$$

for $a_1,\ldots,a_n$ in F implies $a_1=\cdots=a_n=0$. A set $\{v_1,\ldots,v_n\}$ is called a *basis* of V if all its elements $v_1,\ldots,v_n$ in V are linearly independent over F and all elements in V may be expressed in the form $a_1v_1+\cdots+a_nv_n$ where all a_i, $i=1,\ldots,n$, are in F. Also V is said to be of *dimension* n over F, $\dim V=n$. A map $f:V\to W$ from one vector space to another, where V and W are vector spaces over the same field F, is called an *isomorphism* if the map f one to one and onto and, for all v, v_1 and v_2 in V and a in F, $f\left(v_1+v_2\right)=f\left(v_1\right)+f\left(v_2\right)$ and $f(av)=af(v)$.

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –28–
From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 36 *isomorphic vector spaces*

Let two vector spaces V and W over the same field F have the same finite dimension. Then V and W are isomorphic.

Proof. Let $\dim V = \dim W = n$. Let $\{x_1, \ldots, x_n\}$ be a basis of V over F, and $\{y_1, \ldots, y_n\}$ a basis of W over F.
Since all the elements of V can be uniquely written as $a_1x_1 + \cdots + a_nx_n$ for some a_i in F, the map $f : V \to W$, which is

$$f(a_1x_1 + \cdots + a_nx_n) = a_1y_1 + \cdots + a_ny_n$$

for a_i in F, is well defined. Thus f is a homomorphism.
Since $f(a_1x_1 + \cdots + a_nx_n)$ implies $a_1y_1 + \cdots + a_ny_n = 0$ implies $a_1 = \cdots = a_n = 0$, which in turn implies $a_1x_1 + \cdots + a_nx_n = 0$, therefore f is one to one. Then, since all elements of W is of the form $a_1y_1 + \cdots + a_ny_n$, which is equal to $f(a_1x_1 + \cdots + a_nx_n)$ for some $a_1, \ldots, a_n$ in F, therefore f is also onto. Hence f is an isomorphism. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 63 *polynomial codes*

Let

$$g(x) = g_0 + \cdots + g_kx^k$$

be a polynomial in $F[x]$. We call the *polynomial code* with encoding or generating polynomial $g(x)$ a code which encodes each original word of the message $a = (a_0, \ldots, a_{b-1})$, corresponding to

$$a(x) = a_0 + \cdots + a_{b-1}x^{b-1}$$

into the code word $b = (b_0, \ldots, b_{b+k-1})$, which corresponds to the code polynomial

$$b(x) = b_0 + \cdots + b_{b+k-1}x^{b+k-1} = a(x)g(x)$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Note 7 *assumptions for polynomial codes*

We assume for our generating polynomial that $g_0 \neq 0$ and $g_k \neq 0$. To justify this assumption, suppose we have

$$g(x) = g_0 + \cdots + g_k x^k$$

If $g_0 = 0$, then we choose a new polynomial for$g(x)$ as

$$g_1(x) = a_1 + \cdots + a_k x^{k-1}$$

If $g_k = 0$, then we choose another polynomial

$$g_2(x) = g_0 + \cdots + a_{k-1} x^{k-1}$$

In either case our choice becomes more economical.

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –31–
From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 37 *divisible polynomial*

A polynomial with coefficients in **B** is divisible by $1+x$ if and only if it has an even number of terms.

Proof. Let $f(x) = a_0 + \cdots + a_n x^n$ for all a_i in **B**, $i = 1, \ldots, n$, and let $1+x|f(x)$. Then there exists a polynomial $b(x)$ in **B** such that

$$f(x) \equiv (1+x)b(x)$$

If $x = 1$, we have $a_0 + \cdots + a_n = 0$. Since the field **B** is of characteristic 2, this is only possible if the number of non-zero terms is even.

Conversely, let $f(x)$ have an even number of non-zero terms, say $f(x) = x^{i_1} + \cdots + x^{i_{2k}}$, where $i_1 < \cdots < i_{2k}$. Rewrite this as

$$f(x) = \left(x^{i_1} + x^{i_2}\right) + \cdots + \left(x^{i_{2k-1}} + x^{i_{2k}}\right)$$

For $i < j$, $x^i + x^j = x^i\left(1 + x^{j-i}\right) = x^i(1+x)\left(1 + \cdots + x^{j-i-1}\right)$, which means that $1+x|x^i + x^j$. Therefore $1+x$ divides all bracketed terms in $f(x)$, and hence $1+x|f(x)$. ¶

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –32–
From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 38 *minimum distance of a polynomial code*

If $g(x) \in \mathbf{B}[x]$ divides no polynomials of the form $x^k - 1$ for $k < n$, then the binary polynomial code of length n generated by $g(x)$ has the minimum distance of at least 3.

Proof. Let $g(x) = g_0 + \cdots + g_r x^r$, where g_i are in $\mathbf{B}$, $g_0 \neq 0$ and $g_r \neq 0$. Let $b = n - r$. Suppose the opposite to what the theorem says is true. Then, polynomial code being a group code, there exists $b(x)$ with at most two non-zero entries. There are two cases to consider, namely $b(x) = x^i + x^j$, where $i < j$, and $b(x) = x^i$, where $i < n$. In the first one of these, since n is the code length, we have $j < n$, hence $0 < j - i < n$. Since $g(x)|b(x)$ implies $g(x)|x^j \left(1 + x^{j-i}\right)$, and $g_0 \neq 0$ implies $x \not| g(x)$, therefore $g(x)|1 + x^{j-i}$ which contradicts our hypothesis. In the second case, similarly to the above $g(x)|x^i$ and we again have a contradiction. ¶

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –33–
From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 64 *matrix code*

Let C be a (b, n)-code. If there exists a $b \times n$ matrix G of rank b such that

$$C = \left\{\mathbf{a}G | a \in \mathbf{B}^b\right\}$$

then G is called a *generator matrix* of the code C, and C is called a *matrix code* generated by G.

Definition 65 *parity check matrix*

Let C be a (b, n)-code. If there exists an $(n - b) \times n$ matrix H of rank $n - b$ such that

$$H\mathbf{b}^9 0 = 0$$

for all $\mathbf{b}$ in C, then H is called a *parity check matrix* of C.

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –34–
From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 39 *polynomial code a matrix code*

A polynomial code is a matrix code.

Proof. Let C be a polynomial b, n-code with the encoding polynomial $g(x) = g_0 + \cdots + g_k x^k$. Then $n = b + k$. Let G be the $b \times n$ matrix whose first row begins with entries $g_0, \ldots, g_k$ followed by b zeros, and whose succeeding row is an anticlockwise cyclic shift of the previous one, that is

$$G = \begin{bmatrix} g_0 & g_1 & \cdots & g_k & 0 & \cdots & 0 \\ 0 & g_0 & & \cdots & g_k & & \\ \vdots & & & & & & \\ 0 & & \cdots & & g_0 & \cdots & g_k \end{bmatrix}$$

The determinant of the submatrix formed by the first b columns is non-zero, since $g_0 \neq 0$ and hence $g_0^b \neq 0$. Thus the rank of G is m. Let the original word to be coded be $a = (a_0, \ldots, a_{m-1})$. Then, since the code word generated by aG is the same as that generated from $a(x)g(x)$, the two codes are identical. ¶

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –35– From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Algorithm 4 *Hamming codes*

choose r a positive integer
$b \leftarrow 2^r - r - 1$
$n \leftarrow 2^r - 1$
for $i = 1$ to $2^r - 1$ **do**
  (the i^{th} row of M)$\leftarrow (\mathbf{b}_i)$
endfor
for $i = 1$ to $2^r - 1$ **do**
  $(a_1, \ldots, a_{2^r-1}) \leftarrow (\mathbf{b}_i)$
  $(b_{2^{2-1}}, \ldots, b_{2^{r-2}} - 1, b_{2^{r-2}} + 1, \ldots, b_{2^{r-1}} - 1) \leftarrow (a_1, \ldots, a_{2^r-1})$
$(b_{2^{j-1}}; j = 1, \ldots, r) \leftarrow$ **solve** $(\mathbf{b}M = 0)$
the i^{th} code word $\leftarrow (b_1, \ldots, b_n)$
endfor

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –36– From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Note 8 *Hamming codes*

Each code word in a Hamming code contains

$$b - n = 2^r - r - 1 - 2^r + 1 = r$$

check digits. The value of r is called the

redundancy

of the code.

Coding theory, Group, polynomial, and Hamming codes, 25^{th} November 2005 –37–
From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 66 *coset*

Let G be a group. Then a *coset* is a subgroup H of G which is either a *left coset* of H, that is $xH = \{xh : h \in H\}$ for some x in G, or a *right coset* $Hx = \{hx : h \in H\}$ of the same.

Definition 67 *lcm*

Let polynomials $f_1(x), \ldots, f_r(x)$ in $\mathbf{F}_q[x]$ be non-zero. Then the *least common multiple* $\operatorname{lcm}(f_1(x), \ldots, f_r(x))$ of $f_1(x), \ldots, f_r(x)$ is the monic polynomial of the lowest degree which is a multiple of all $f_i(x)$, $i = 1, \ldots, r$.

Problem 6

Prove that for non-zero polynomials $f_1(x), \ldots, f_r(x)$ in $\mathbf{F}_q[x]$,

$$\operatorname{lcm}(f_1(x), \ldots, f_r(x)) = \operatorname{lcm}(\operatorname{lcm}(f_1(x), \ldots, f_{r-1}(x)), f_r(x))$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Note 9

Let $f_1(x), \ldots, f_r(x)$ in $\mathbf{F}_q[x]$ have the factorisations,

$$\begin{aligned} f_1(x) &= a_1 (p_1(x))^{e_{11}} \cdots (p_n(x))^{e_{1n}} \\ &\vdots \\ f_r(x) &= a_r (p_1(x))^{e_{r1}} \cdots (p_n(x))^{e_{rn}} \end{aligned}$$

where $a_1, \ldots, a_r$ are in $\mathbf{F}_q^*$, $e_{ij} \geq 0$, and $p_i(x)$ are distinct monic irreducible polynomials over $\mathbf{F}_q$, then

$$\operatorname{lcm}(f_1(x), \ldots, f_r(x)) = (p_1(x))^{\max(e_{11}, \ldots, e_{r1})} \cdots (p_n(x))^{\max(e_{1n}, \ldots, e_{rn})}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 40

Let $f(x), f_1(x), \ldots, f_r(x)$ be polynomials over $\mathbf{F}_q$. If $f(x)$ is divisible by every polynomial f_i, for $i = 1, \ldots, r$, then $f(x)$ is also divisible by $\operatorname{lcm}(f_1(x), \ldots, f_r(x))$.

Proof. Consider first the case where there are only two different polynomials, $f_1(x)$ and $f_2(x)$. The prime components of $f_1(x)$ and $f_2(x)$ may be grouped into those which are unique among them and those which are shared. Since

$$f(x) = u_1(x)f_1(x) + r_1(x)$$

and

$$f(x) = u_2(x)f_2(x) + r_2(x)$$

it follows that $f(x)$ contains both of these two groups of primes. In other words,

$$f(x) = u(x)\operatorname{lcm}(f_1(x), f_2(x)) + r(x)$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Next, consider the case where there are more than two f_i's. Suppose for $f(x)$, that

$$f(x) = u_r(x)\operatorname{lcm}(f_1(x), \ldots, f_r(x))$$

Then if we let

$$f_c(x) = \operatorname{lcm}(f_1(x), \ldots, f_r(x))$$

and if we introduce another polynomial $f_{r+1}(x)$ such that

$$f(x) = u_{r+1}f_{r+1} + r_{r+1}(x)$$

then following the same line of reasoning as the above we have,

$$\operatorname{lcm}(f_1(x), \ldots, f_{r+1}(x)) | f(x)$$

¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 68 *subring*

A non-empty subset S of a ring R is called a *subring* of R if the elements of S form a ring with respect to the operations defined in R.

Theorem 41 *subring*

Let R be a ring. Then a non-empty subset S of R is a subring if and only if S is closed under addition, multiplication, and the formation of additive inverse.

Proof. Since S is a subset of R, additive associativity, identity and commutativity are inherited to S from R. The existence of the inverse for each element s in S is certain provided that the formation of an additive inverse is guaranteed. And similarly in the case of multiplication, both associativeness and distributiveness hold once we know that S is closed under multiplication. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 69 *ideal of a ring*

Let R be a ring. We call an *ideal* in R a subring I having such property that for all i in I, then both xi and ix are also in I for every element x in R. Further, if I is a proper subset of R, then it is called a *proper ideal*. By *trivial ideal* one means either the *zero ideal* $\{0\}$ consisting of the zero element alone, or the ring R itself.

Note 10

The significance of the ideals in a ring is that they let us construct other rings from the first. The cosets of a ring R is a partition of R into equivalence sets, which are non-empty and disjoint, the union of which is the whole of the ring R.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 70 *congruence of a ring*

Let R be a ring and I an ideal in it. Then two elements x and y in R are said to be *congruent modulo* I, denoted by

$$x \equiv y (\mathrm{mod}\, I)$$

if $x - y$ is in I. Since there is only ideal, we may a write this congruence as simply $x \equiv y$.

Note 12 *addition and multiplication of congruences*

Congruences can be added and multiplied as if they were ordinary equations. In other words, if $x_1 \equiv x_2$ and $y_1 \equiv y_2$, then

$$x_1 + y_1 \equiv x_2 + y_2$$

and

$$x_1 y_1 \equiv x_2 y_2$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 71 *coset of a ring*

Let R be a ring and let x be an element of R. Then the *coset* $[x]$ containing x is the set of all elements y such that $y \equiv x$. Then,

$$\begin{aligned}[x] = \{y : y \equiv x\} &= \{y : y - x \in I\} \\ &= \{y : y - x = i \text{ for some } i \in I\} \\ &= \{y : y = x + i \text{ for some } i \in I\} \\ &= \{x + i : i \in I\} = x + I\end{aligned}$$

Furthermore, $[x] = [x_1]$ means that $x \equiv x_1$, that is to say, $x - x_1$ is in I. Here x and x_1 are called *representatives* of the coset which contains them.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 72 *quotient ring*

A *quotient ring*, aka *residue-class-, factor-*, or *difference ring*, is a ring having the form of a quotient A/i of a ring A and one of its ideal i. In other words, the quotient ring of R with respect to I the ring

$$R/I = \{x + I : x \in R\}$$

where

$$x + I = \{x + i : i \in I\}$$

is the coset of an element x in R, and where addition and multiplication are defined as,

$$[x] + [y] = [x + y]$$

and

$$[x] \cdot [y] = [xy]$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 42

The zero element of R/I is $0 + I = I$, the negative of $x + I$ is $(-x) + I$. If R is commutative, then R/I is also commutative. If R has an identity 1 and a proper ideal I, then R/I has an identity $1 + I$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 43 *quotient ring*

Let R be a ring and I an ideal of R. Then, for x and y in R,

$$(x+I)+(y+I)=(x+y)+I$$

and

$$(x+I)(y+I)=xy+I$$

Proof. Let a and b be any two elements of the ideal I. Then,

$$(x+a)+(y+b)=x+a+y+b=(x+y)+(a+b)=(x+y)+p$$

where $p=a+b$ is in I. Further,

$$\begin{aligned}(x+a)(y+b)&=xy+bx+ay+ab\\&=xy+c+d+e=xy+f\end{aligned}$$

where $c=bx$, $d=ay$, $e=ab$ and $f=c+d+e$ are all elements of I. ¶

Coding theory, Finite field- and BCH codes, 2^{nd} December 2005 –11– From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Note 13

Theorem 43 and Note 12 show that the quotient ring R/I defined in Definitions 72 is independent of the choice of x and y in the cosets $x+I$ and $y+I$. In other words, the cosets $[x+y]$ and $[xy]$ resulted from addition and respectively multiplication in no ways depend on the particular representatives x and y chosen for the cosets $[x]$ and $[y]$ that go into them. This means that, if $x_1\equiv x$ and $y_1\equiv y$, then

$$[x_1+y_1]=[x+y]$$

and

$$[x_1y_1]=[xy]$$

or equivalently

$$x_1+y_1\equiv x+y \quad \text{and} \quad x_1y_1\equiv xy$$

Coding theory, Finite field- and BCH codes, 2^{nd} December 2005 –12– From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Example 31

Some examples of quotient ring are $\mathbf{Z}_2 = \mathbf{Z}/2\mathbf{Z}$ and $\mathbf{Z}_6 = \mathbf{Z}/6\mathbf{Z}$.

Theorem 44 *polynomial ring*

The polynomial ring $F[x]$ is a commutative ring with identity.

Proof. $F[x]$ is a ring over the field F since under addition it is closed, associative and commutative, and has 0 as the identity and the inverse $-f(x)$, where $f(x) \in F[x]$; and under multiplication it is associative, distributive and commutative, and has 1 as the identity. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 73 *principal ideal*

Let R be a commutative ring with identity. Then for any a in R the *principal ideal* generated by a is

$$\langle a \rangle = aR = \{ar : r \in R\}$$

Further, R is called *principal ideal ring* if all its ideals are of this form.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 45 *polynomial ring a principal ideal ring*

Let F be a field. Then the polynomial ring $F[x]$ is a principal ideal ring.

Proof. The polynomial ring $F[x]$ being a commutative ring with identity, it remains only to show that all its ideals are of the form

$$\langle a \rangle \; R = aR = \{ar : r \in R\}$$

where a is in R. Let I be an ideal of $F[x]$. If $I = 0$, then I is a principal ideal generated by 0. If $I \neq 0$, then choose $0 \neq f(x) \in I$ such that

$$\deg f \leq \deg g$$

for all non-zero $g(x)$ in I. Write

$$g(x) = q(x)f(x) + r(x)$$

If $\deg g < \deg f$, then $q = 0$ and $r = f$. On the other hand, if $n = \deg f \leq \deg g$, then either r is 0 or $\deg r < \deg f$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Let

$$f(x) = a_0x^n + \cdots + a_n \quad \text{and} \quad g(x) = b_0x^m + \cdots + b_m$$

Then, with $a_0 \neq 0$,

$$g(x) = a_0^{-1}b_0x^{m-n}f(x) + g_1(x) \tag{23}$$

where $\deg g_1 \leq m - 1$. Then

$$g_1(x) = q_1(x)f(x) + r(x) \tag{24}$$

From this it follows that either $r = 0$ or $\deg r < \deg f$. From Equation's 23 and 24,

$$g(x) = q(x)f(x) + r(x) \quad \text{where} \quad q(x) = a_0^{-1}b_0x^{m-n} + q_1$$

is in $F[x]$. If $r \neq 0$, then $r(x)$ is in I and $\deg r < \deg f$, which contradicts our choice of $f(x)$. Therefore $g = qf$ and I is a principal ideal generated by $f(x)$. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 74 *reducible polynomial*

Let R be a commutative ring with identity. Then a non-constant $f(x)$ in $R[x]$ is said to be *reducible* if, for some $g(x)$ and $h(x)$ in $R[x]$,

$$f(x) = g(x)h(x)$$

implies either $\deg g(x) = 0$ or $\deg h(x) = 0$. Otherwise $f(x)$ is said to be *reducible.*

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 46 *quotient ring a field*

Let F be a field $f(x)$ in $F[x]$ an irreducible polynomial. Then $F[x]/\langle f(x)\rangle$ is a field.

Proof. Let I be the ideal $\langle f(x)\rangle$ of $F[x]$ generated by $f(x)$. If $I = F[x]$, then $f(x)$ has an inverse, that is $1 = f(x)g(x)$ for some $g(x)$ in $F[x]$. Then $f(x)$ is a constant polynomial, which contradicts our statement of the theorem. Therefore $F[x]/I$ has at least two elements, and $F[x]/I$ being a polynomial ring it is a commutative ring with identity.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Let $g \in F[x]$ and $g \notin I$. Then,

$$J = \{a(x)f(x) + b(x)g(x) : a(x), b(x) \in F[x]\}$$

is an ideal of $F[x]$ and there exists $h(x)$ in $F[x]$ such that $J = \langle h(x) \rangle$. But

$$f(x) = 1f(x) + 0g(x)$$

is in J, and thus $f(x) = a(x)h(x)$ for some $a(x)$ in $F[x]$. The polynomial $f(x)$ being irreducible, either $\deg h(x) = 0$ or $\deg a(x) = 0$. If the latter is the case, then $a(x)$ is a unit in $F[x]$, and then $h(x)$ is in I, hence $J = I$, and hence a contradiction since we began with g being in J but not in I. Therefore it must be the case that $h(x)$ is a unit in $F[x]$, hence J is a unit, and thus

$$1 = a(x)f(x) + b(x)g(x)$$

for some $a(x)$ and $b(x)$ in $F[x]$. And then

$$1 + I = I + b(x)g(x) = (b(x) + I)(g(x) + I)$$

Thus $g(x) + I$ has an inverse and $F[x]/I$ is a field. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 75 *extension of a field*

Let K be a field and F a subfield of K. Then K is called an *extension* of the field F, denoted by

$$K|_F$$

Since K has multiplication, it is a vector space over F. The dimension of the vector space K over F is called the *degree*

$$[K : F]$$

of the extension K of F. The extension $K|_F$ is said to be *finite* if the degree $[K : F]$ is finite.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 76 *prime subfield*

A *prime subfield* of a field F is the intersection of all subfields of F. It is the smallest of all subfields of F, and is unique. A *prime field* is a field which has no proper subfields.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 77 *minimal polynomial*

Let $K|_F$ be an extension of a field F. Then $\alpha \in K$ is said to be *algebraic* over F if there exists $f(x)$ in $F[x]$ which has α as a root. Let α in K be algebraic over F and consider

$$A = \{f(x) \in F[x] : f(\alpha) = 0\}$$

Here A is an ideal of the principal ideal domain $F[x]$. Let $m_1(x)$ in $F[x]$ be a generator of A. If a is the coefficient of the highest power of x in $m_1(x)$, then $m(x) = a^{-1}m_1(x)$ is a monic polynomial with $\deg m(x) = \deg m_1(x)$, and $m(x)$ is also a generator of A. Let $m(x) = r(x)s(x)$ for some $r(x)$ and $s(x)$ in $F[x]$. Then either $r(\alpha) = 0$ or $s(\alpha) = 0$, that is either $m(x)|r(x)$ or $m(x)|s(x)$. But $\deg m = \deg r + \deg s$, therefore either $\deg r(x) = 0$ or $\deg s(x) = 0$. Hence $m(x)$ is irreducible. Since $m(x)$ is monic, irreducible and is of the least degree possible while admitting α as a root, therefore $m(x)$ is called the *minimal polynomial* of α over $F[x]$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 47 *linear code correction capability*

Let C be an (n, k) linear code over F_q with prith-check matrix H, and $d(C)$ the smallest number of column of H that are linearly dependent. Then if every subset of $2t$ or fewer columns of H is linearly independent, the code is capable of correcting all error patterns of weight $w \leq t$.

Proof. When $q = 2$, linear independence amounts to summing to $\mathbf{0}$. The code words of C are those vectors $\mathbf{x}$ in $V_n(F_q)$ for which

$$H\mathbf{x}^T = \mathbf{0}$$

But $H\mathbf{x}^T$ is a linear combination of the columns of H, that is to say, if $H = [\,\mathbf{c}_1 \quad \cdots \quad \mathbf{c}_n\,]$ then $H\mathbf{x}^T = x_1\mathbf{c}_1 + \cdots + x_n\mathbf{c}_n$. Hence a non-zero code word of weight w gives a nontrivial linear dependence among w columns of H, and vice versa. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Corollary 47[1] *Hamming as special case*

If $q = 2$ and all possible linear combinations of up to e columns are distinct, then

$$d(C) \geq 2e + 1$$

and C can then correct all patterns of weight e or less.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Note 14 *from Hamming to BCH code*

Hamming codes correct single errors. An extension of this is to the Bose-Chaudhuri-Hocquenghem codes which could correct multiple errors. In the case of Hamming code of length $n = 2^m - 1$, the parity-check matrix is given by

$$H = [\mathbf{v}_0 \quad \dots \quad \mathbf{v}_{n-1}]$$

where $(\mathbf{v}_0 \quad \cdots \quad \mathbf{v}_{n-1})$ is some ordering of the $2^m - 1$ non-zero column vectors in $V_m = V_m(F_2)$. The $m \times n$ matrix H takes m parity-check bits for the code to be able to correct one error. We may extend H such that it has m more rows and could correct two errors. Then,

$$H_2 = \begin{bmatrix} \mathbf{v}_0 & \cdots & \mathbf{v}_{n-1} \\ \mathbf{w}_0 & \cdots & \mathbf{w}_{n-1} \end{bmatrix}$$

where $\mathbf{w}_0, \dots, \mathbf{w}_{n-1}$ are in V_m.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Since $\mathbf{v}_i$'s are distinct, we may look at the mapping from $\mathbf{v}_i$ to $\mathbf{w}_i$ as a function from V_m into itself, then

$$H_2 = \begin{bmatrix} \mathbf{v}_0 & \cdots & \mathbf{v}_{n-1} \\ \mathbf{f}(\mathbf{v}_0) & \cdots & \mathbf{f}(\mathbf{v}_{n-1}) \end{bmatrix}$$

Then H_2 will define a code which corrects two errors if and only if the syndromes of the $1 + n + \binom{n}{2}$ error patterns of weights 0, 1 and 2 are all distinct. Any such syndrome is a sum of a subset of columns of H_2, and therefore a vector in V_{2m}. Let the syndrome be $\mathbf{s} = (s_1 \quad \dots \quad s_{2m}) = (\mathbf{s}_1 \quad \mathbf{s}_2)$, where $\mathbf{s}_1 = (s_1, \dots, s_m)$ and $\mathbf{s}_2 = (s_{m+1}, \dots, s_{2m})$ are both in V_m. Defining $\mathbf{f}(\mathbf{0}, \mathbf{0}) = \mathbf{0}$ we consider a pair of errors occuring at i^{th}- and j^{th} position's, $\mathbf{s} = (\mathbf{v}_i + \mathbf{v}_j, \mathbf{f}(\mathbf{v}_i) + \mathbf{f}(\mathbf{v}_j))$. Then the system of equations, $\mathbf{u} + \mathbf{v} = \mathbf{s}_1$ and $\mathbf{f}(\mathbf{u}) + \mathbf{f}(\mathbf{v}) = \mathbf{s}_2$. has at most one solution $(\mathbf{u}, \mathbf{v})$ for each pair of vectors from V_m.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

By trial and error we may find neither the linear mapping $\mathbf{f}(\mathbf{v}) = T\mathbf{v}$ nor the nonlinear polynomial of degree 2 works, but $\mathbf{f}(\mathbf{v}) = \mathbf{v}^3$ does. The matrix

$$H_2 = \begin{bmatrix} \alpha_0 & \cdots & \alpha_{n-1} \\ \alpha_0^3 & \cdots & \alpha_{n-1}^3 \end{bmatrix}$$

is the parity-check matrix of a binary code of length $n = 2^m - 1$ which corrects up to two errors. A vector $\mathbf{c} = (\, c_0 \quad \cdots \quad c_{n-1} \,)$ in $V_n(F_2)$ is a code word in the code defined by H_2 if and only if

$$\sum_{i=0}^{n} c_i \alpha_i = \sum_{i=0}^{n} c_i \alpha_i^3 = 0$$

Since the $2m$ rows of the matrix H_2 over F_2 may not be all linearly independent, the dimension of the code is

$$d(C) \geq n - 2m = 2^m - 1 - 2m$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 78 *Vandermonde matrix*

The *Vandermonde matrix* is defined as

$$A = \begin{bmatrix} 1 & \cdots & 1 \\ a_1 & \cdots & a_r \\ \vdots & \ddots & \vdots \\ a_1^{r-1} & \cdots & a_r^{r-1} \end{bmatrix}$$

Theorem 48 *Vandermonde matrix*

Let $a_1, \ldots, a_r$ be distinct non-zero elements of a field. Then the the Vandermonde matrix is such that

$$\begin{vmatrix} 1 & \cdots & 1 \\ a_1 & \cdots & a_r \\ \vdots & \ddots & \vdots \\ a_1^{r-1} & \cdots & a_r^{r-1} \end{vmatrix} \neq 0$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proof. Subtracting row$(i+1) - a_1$ row i, $i = 1, \cdots, r-1$, yields,

$$\det A = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ 0 & a_2 - a_1 & \cdots & a_r - a_1 \\ 0 & a_2(a_2 - a_1) & \cdots & a_r(a_r - a_1) \\ \vdots & & \ddots & \vdots \\ 0 & a_2^{r-2}(a_2 - a_1) & \cdots & a_r^{r-2}(a_r - a_1) \end{vmatrix}$$

$$= (a_2 - a_1) \cdots (a_r - a_1) \begin{vmatrix} 1 & \cdots & 1 \\ a_2 & \cdots & a_r \\ \vdots & \ddots & \vdots \\ a_2^{r-2} & \cdots & a_r^{r-2} \end{vmatrix}$$

$$= (a_2 - a_1) \cdots (a_r - a_1) \cdot (a_3 - a_2) \cdots (a_r - a_2) \begin{vmatrix} 1 & \cdots & 1 \\ a_3 & \cdots & a_r \\ \vdots & \ddots & \vdots \\ a_3^{r-3} & \cdots & a_r^{r-3} \end{vmatrix}$$

$\cdots = \prod_{i>j}(a_i - a_j)$. Then, since a_i are distinct and non-zero, therefore $\det A$ is non-zero. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 49 *linearly independence*

Any square matrix having a non-zero determinant has all its columns linearly independent.

Proof. Let A be an $r \times r$ matrix, and that $|A| \neq 0$. Then suppose the columns of A are linearly dependent. Then one may write some column of A as a linear combination of the others, for example

$$\mathbf{c}_j = \sum_{\substack{i=1 \\ i \neq j}}^{r} a_i \mathbf{c}_i$$

Then if column $\mathbf{c}_j$ is replaced by $\mathbf{c}_j - \sum_{\substack{i=1 \\ i \neq j}}^{r} a_i \mathbf{c}_i$ gives a matrix B with $|B| = |A|$. But B also has a column whose all elements are zeros, which means that

$$|A| = |B| = 0$$

a contradiction and thus the proof. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 50 *BCH code*

Let $(\alpha_0, \ldots, \alpha_{n-1})$ be an ordering of non-zero elements of $\mathbf{F}_{2^m}$, and let t be a positive integer such that

$$t \leq 2^{m-1} - 1$$

Then the matrix

$$H = \begin{bmatrix} \alpha_0 & \cdots & \alpha_{n-1} \\ \alpha_0^3 & \cdots & \alpha_{n-1}^3 \\ \vdots & \ddots & \vdots \\ \alpha_0^{2t-1} & \cdots & \alpha_{n-1}^{2t-1} \end{bmatrix}$$

is the parity-check matrix of a binary (n, k)-code capable of correcting all error patterns of weight $w \leq t$, with dimension

$$k \geq n - mt$$

Coding theory, Finite field- and BCH codes, 2^{nd} December 2005 –31– From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proof. A vector $\mathbf{c} = (c_0, \ldots, c_{n-1})$ in $V_n(F_2)$ is a code word if and only if $H\mathbf{c}^T = \mathbf{0}$. Thus,

$$\sum_{i=0}^{n-1} c_i \alpha_i^j = 0$$

for $j = 1, 3, \ldots, 2t-1$. We simplify this by using the fact that $(x+y)^2 = x^2 + y^2$ in characteristic 2, and $x^2 = x$ in F_2. Hence,

$$\left(\sum_{i=0}^{n-1} c_i \alpha_i^j \right)^2 = \sum_{i=0}^{n-1} c_i^2 \alpha_i^{2j} = \sum_{i=0}^{n-1} c_i \alpha_i^{2j}$$

for $j = 1, 3, \ldots, 2t-1$, which gives us

$$\sum_{i=0}^{n-1} c_i \alpha_i^j$$

for $j = 1, 2, \ldots, 2t$.

Coding theory, Finite field- and BCH codes, 2^{nd} December 2005 –32– From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Therefore we could also use the parity-check matrix

$$H^90 = \begin{bmatrix} \alpha_0 & \cdots & \alpha_{n-1} \\ \alpha_0^2 & \cdots & \alpha_{n-1}^3 \\ \vdots & \ddots & \vdots \\ \alpha_0^{2t} & \cdots & \alpha_{n-1}^{2t} \end{bmatrix}$$

According to Theorem 47 H^90 is a parity-check matrix which corrects t errors if and only if every subset of $2t$ or fewer columns of H^90 is linearly independent. Next, since a subset of $r \leq 2t$ columns of H^90 has the form

$$A = \begin{bmatrix} a_1 & \cdots & a_r \\ a_1^2 & \cdots & a_r^2 \\ \vdots & \ddots & \vdots \\ a_1^{2t} & \cdots & a_r^{2t} \end{bmatrix}$$

where $a_1, \ldots, a_r$ are distinct non-zero elements of F_{2m}, we may consider the matrix

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

$$A^90 = \begin{bmatrix} a_1 & \cdots & a_r \\ \vdots & \ddots & \vdots \\ a_1^r & \cdots & a_r^r \end{bmatrix}$$

which is nonsingular since its determinant by the Vandermonde determinant theorem, Theorem 48, is

$$\det A^90 = a_1 \cdots a_r \begin{vmatrix} 1 & \cdots & 1 \\ a_1 & \cdots & a_r \\ \vdots & \ddots & \vdots \\ a_1^{r-1} & \cdots & a_r^{r-1} \end{vmatrix} = a_1 \cdots a_r \prod_{i<j} (a_j - a_i) \neq 0$$

Then the columns of A^90, and hence those of A, cannot be linearly dependent, and therefore the code corrects all error patterns of weight up to t. Now H, as a matrix with entries from F_2 rather than F_{2m}, has dimensions $mt \times n$, hence the dual code has dimension $k \leq mt$, and the code has dimension $k \geq n - mt$. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 51 *minimum distance of linear code*

Let C be a linear (n,k)-code over $GF(q)$ with parity-check matrix H. Then the minimum distance of C is d if and only if any $d-1$ columns of H are linearly independent but some d columns are linearly dependent.

Proof. The minimum distance of a code $d(C)$ is equal to the minimum of the weights of the non-zero code words. Let $\mathbf{x} = x_1 \cdots x_n$ be a vector in $V(n,q)$. Then $\mathbf{x}$ is in C if and only if $\mathbf{x}H^T = \mathbf{0}$ if and only if $x_1\mathbf{h}_1 + \cdots + x_n\mathbf{h}_n = \mathbf{0}$, where $\mathbf{h}_1, \ldots, \mathbf{h}_n$ are the columns of H. Therefore there is a set of d linearly dependent columns of H corresponding to each code word $\mathbf{x}$ of weight d. On the other hand, if there existed a set of $d-1$ linearly dependent columns of H, then there would exist some scalars $x_{i_1}, \ldots, x_{i_{d-1}}$, not all zero, such that $\sum_{j=1}^{d-1} x_{i_j} = \mathbf{0}$. But if this were the case, then $\mathbf{x}H^T = \mathbf{0}$ and so would be a code word of weight $0 < d < d(C)$. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 52 *Singleton bound*

The maximum dictionary size m such that there exists a q-ary (n,m,d)-code is

$$A_q(n,d) \leq q^{n-d+1}$$

Proof. Let C be a q-ary (n,m,d)-code. If we remove the last $d-1$ coordinates from each code word, then the m vectors of length $n-d+1$ so obtained must be distinct, otherwise $d(C)$ must be less than d, which would contradict the statement above. Therefore $m \leq q^{n-d+1}$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 53 *bound for BCH codes*

Let C be the code over $GF(q)$, where q is a prime number, and C is defined to have the parity-check matrix

$$H = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 3 & \cdots & n \\ 1 & 2^2 & 3^2 & \cdots & n^2 \\ \vdots & & & \ddots & \vdots \\ 1 & 2^{d-2} & 3^{d-2} & \cdots & n^{d-2} \end{bmatrix}$$

where $d \leq n \leq q-1$. If q is a prime-power, then

$$A_q(n,d) = q^{n-d+1}$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proof. We have,

$$C = \left\{ x_1 \cdots x_n \in V(n,q) \text{ s.t. } \sum_{i=1}^{n} i^j x_i = 0 \text{ for } j = 0, 1, \ldots, d-2 \right\}$$

Any $d-1$ columns form a Vandermonde matrix, and therefore by Theorem's 48 and 49 are linearly independent. By Theorem 51 C has a minimum distance d and therefore is a q-ary (n, q^{n-d+1}, d)-code. The proof follows since C meets the Singleton bound of Theorem 52. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Problem 9 *decoding BCH*

Find the decoding procedure for the BCH codes.

Solution. Assume that $d = 2t + 1$ and H has $2t$ rows. Suppose the code word $\mathbf{c} = c_1 \cdots c_n$ is transmitted and the vector $\mathbf{r} = r_1 \cdots r_n$ is received. Assuming that at most t errors have occurred, let $x_1, \ldots, x_t$ be their positions and $m_1, \ldots, m_t$ their respective magnitudes. Then the syndrome is

$$(s_1, \ldots, s_{2t}) = \mathbf{r}H^T$$

and we have

$$s_j = \sum_{i=1}^{n} r_i i^{j-1} = \sum_{i=1}^{t} m_i x_i^{j-1} \tag{25}$$

for $j = 1, \ldots, 2t$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Then from

$$\phi(\theta) = \frac{m_1}{1 - x_1\theta} + \frac{m_2}{1 - x_2\theta} + \cdots + \frac{m_t}{1 - x_t\theta} \tag{26}$$

and

$$\frac{m_i}{1 - x_i\theta} = m_i \left(1 + x_i\theta + x_i^2\theta^2 \cdots\right)$$

together with Equation 25, we have

$$\phi(\theta) = s_1 + s_2\theta + \cdots + s_{2t}\theta^{2t-1} + \cdots$$

Also, from Equation 26 we have

$$\phi(\theta) = \frac{a_1 + a_2\theta + a_3\theta^2 + \cdots + a_t\theta^{t-1}}{1 + b_1\theta + b_2\theta^2 + \cdots + b^t\theta^t} \tag{27}$$

Hence,

$$\left(s_1 + s_2\theta + s_3\theta^2 + \cdots\right)\left(1 + b_1\theta + b_2\theta^2 + \cdots + b_t\theta^t\right) = a_1 + a_2\theta + \cdots + a_t\theta^{t-1}$$

Which gives us

$$a_1 = s_1 \quad \text{and} \quad a_i = \sum_{j=0}^{i-1} s_{i-j} b_i, \quad i = 2, \ldots, t \tag{28}$$

and

$$0 = \sum_{j=0}^{t} s_{i-j} b_j, \quad i = t+1, \ldots, 2t \tag{29}$$

With a_i and b_i known we may turn Equation 27 into partial fractions

$$\phi(\theta) = \frac{p_1}{1 - q_1\theta} + \cdots + \frac{p_t}{1 - q_t\theta}$$

and therefore $m_i = p_i$ and $x_i = q_i$, for $i = 1, \ldots, t$, and the system in Equation 25 is solved. Algorithm 5 then gives the procedure for error correction.

#

Note 15

The polynomial

$$\sigma(\theta) = 1 + b_1\theta + b_2\theta^2 + \cdots + b_t\theta^t = (1 - x_1\theta) \cdots (1 - x_t\theta) \tag{30}$$

can be used to locate the location of the errors. The polynomial

$$\omega(\theta) = a_1 + a_2\theta + \cdots + a_t\theta^{t-1}$$

can be used to find the magnitude of the errors.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Algorithm 5 *Procedure for correcting up to t errors in BCH codes*

input: **r**
find $s_1, \ldots, s_{2t}$
$e \leftarrow$maximum number of equations in Equation 29
for $i = e + 1$ to t **do**
  $b_i \leftarrow 0$
endfor
$(b_1, \ldots, b_e) \leftarrow$**solve** the first e equations of Equation 29
$(z_1, \ldots, z_e) \leftarrow$**find** the e zeros of Equation 30
$(a_1, \ldots, a_e) \leftarrow$**solve** Equation 28
for $i = 1$ to e **do**
  $m_i \leftarrow \frac{a_1 + a_2 x_i + \cdots + a_e x_i^{e-1}}{\prod_{\substack{j=1 \\ j \neq i}}^{e} (1 + x_j x_i)}$
endfor

Coding theory, Finite field- and BCH codes, 2^{nd} December 2005 –43– From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 79 *linear independence*

Let V be a vector space over $\mathbf{F}_q$. Then a set of vectors $A = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ in V is said to be *linearly independent* if and only if a *linear combination* $\lambda_1\mathbf{v}_1 + \cdots + \lambda_k\mathbf{v}_k$ being a zero-vector implies that λ_i, $i = 1, \ldots, k$, are zero.

Coding theory, Linear codes, 9^{th} December 2005 –1– From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 80 *linear span of a subspace*

Let V be a vector space over $\mathbf{F}_q$. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a non-empty subset of V. Then, the *linear span* $\langle S\rangle$ of S is defined as

$$\langle S\rangle = \left\{ \sum_{i=1}^{k} \lambda_i \mathbf{v}_i : \lambda_i \in \mathbf{F}_q \right\}$$

We say that the span $\langle S\rangle$ of S is a subset of V generated or spanned by S. Let C be a subspace of V, then a subset S of C is called a *generating-* or *spanning set* of C if $C = \langle S\rangle$.

Coding theory, Linear codes, 9^{th} December 2005 –2– From 8^{th} November 2005 , as of 14^{th} January, 2007

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 81 *inner product of vectors*

An *inner product* on $\mathbf{F}_q$ is a mapping $\langle \mathbf{a}, \mathbf{b} \rangle : \mathbf{F}_q^n \times \mathbf{F}_q^n \to \mathbf{F}_q$ such that, for all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in $\mathbf{F}_q^n$,

a. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
b. $\langle \alpha \mathbf{v}, \mathbf{w} \rangle = \alpha \langle \mathbf{v}, \mathbf{w} \rangle$, where α is a scalar
c. $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$
d. $\langle \mathbf{u}, \mathbf{v} \rangle = 0$, for all non-zero $\mathbf{u}$ in $\mathbf{F}_q^n$, if and only if $\mathbf{v} = \mathbf{0}$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 82 *scalar product*

Let $\mathbf{v}$ and $\mathbf{w}$ be two vectors in $\mathbf{F}_q^n$. Then the *scalar product*, aka the *dot-* or *Euclidean inner product*, between $\mathbf{v}$ and $\mathbf{w}$ is defined as $\mathbf{v} \cdot \mathbf{w} = \sum_{i=1}^{n} v_i w_i \in \mathbf{F}_q$. The two vectors are said to be *orthogonal* to each other if and only if $\mathbf{v} \cdot \mathbf{w} = 0$. The *orthogonal complement* $S^{\perp}$ of a non-empty subset S of $\mathbf{F}_q^n$, is defined to be

$$S^{\perp} = \left\{ \mathbf{v} \in \mathbf{F}_q^n : \mathbf{v} \cdot \mathbf{s} = 0 \,\text{for all}\, \mathbf{s} \in S \right\}$$

When $S = \emptyset$ we define $S^{\perp} = \mathbf{F}_q^n$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Note 16 *orthogonal complement*

The orthogonal complement $S^{\perp}$ of a non-empty subset S of a vector space $\mathbf{F}_q^n$ is always a subspace of $\mathbf{F}_q^n$. Moreover, $\langle S\rangle^{\perp} = S^{\perp}$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 83 *basis*

Let V be a vector space over $\mathbf{F}_q$. Then a non-empty subset $A = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of V is called a *basis* for V if $V = \langle A\rangle$ and A is linearly independent.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 54 *dimension of a vector space*
Let V be a vector space over $\mathbf{F}_q$. If $\dim v = k$, then V has q^k elements and

$$\frac{1}{k!}\prod_{i=0}^{k-1}\left(q^k - q^i\right)$$

different bases.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Proof. If the basis for V is $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ and $\lambda_1, \ldots, \lambda_k$ are in $\mathbf{F}_q$, then $V = \sum_{i=1}^{k} \lambda_i \mathbf{v}_i$. Since $|\mathbf{F}_q|$ is q, there are q choices for each λ_i. Therefore V has exactly q^k elements.
Let $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis for V. Since B is non-empty, $\mathbf{v}_1 \neq \mathbf{0}$ and there are $q^k - 1$ choices for $\mathbf{v}_1$. Then there are $q^k - q^{i-1}$ choices of $\mathbf{v}_i$, for $i = 2, \ldots, k$ because $\mathbf{v}_i \notin \langle \mathbf{v}_1, \ldots, \mathbf{v}_{i-1} \rangle$. Therefore there are $\prod_{i=0}^{k-1}\left(q^k - q^i\right)$ distinct ordered k-tuples, $(\mathbf{v}_1, \ldots, \mathbf{v}_k)$. The order of $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is irrelevant, hence the number of distinct bases for V is $\frac{1}{k!}\prod_{i=0}^{k-1}\left(q^k - q^i\right)$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Corollary 54[1] *dimension and size of a code*

Let C be a linear code of length n over $\mathbf{F}_q$. Then, $\dim C = \log_q |C|$, in other words $|C| = q^{\dim C}$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 55 *span and dual*

Let S be a subset of $\mathbf{F}_q^n$. Then, $\dim \langle S \rangle + \dim S^{\perp} = n$.

Proof. When $\langle S \rangle = \{\mathbf{0}\}$, this is obvious. Next, consider cases where $\dim \langle S \rangle = k$, where $1 \leq k < n$. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis of $\langle S \rangle$, then we need to show that $\dim S^{\perp} = \dim \langle S \rangle^{\perp} = n - k$. Since $\mathbf{x}$ is in $S^{\perp}$ if and only if $\mathbf{v}_1 \cdot \mathbf{x} = \cdots = \mathbf{v}_k \cdot \mathbf{x} = 0$, or equivalently $A\mathbf{x} = \mathbf{0}$, where the $k \times n$ matrix A is

$$A = \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_k^T \end{bmatrix}$$

we know that the rows of A are linearly independent. Then $A\mathbf{x} = \mathbf{0}$ is a linear system of k linearly independent equations in n variables, where $n > k$, and therefore admits a solution space of dimension $n - k$. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Corollary 55[1]

Let C be a linear code of length n over $\mathbf{F}_q$. Then $C^\perp$ is also a linear code, and $\dim C + \dim C^\perp = n$

Proof. This follows from Note 16 and Theorem 55 above. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 56 *double orthogonal*

Let C be a linear code of length n over $\mathbf{F}_q$. Then, $\left(C^\perp\right)^\perp = C$.

Proof. From Corollary 55[1], we have $\dim C + \dim C^\perp = n$ and $\dim C^\perp + \dim \left(C^\perp\right)^\perp = n$, and hence $\dim C = \dim \left(C^\perp\right)^\perp$. Let $\mathbf{c}$ be in C. Then for all $\mathbf{x}$ in C, we have $\mathbf{c} \cdot \mathbf{x} = 0$, hence $C \subseteq \left(C^\perp\right)^\perp$ and the proof. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 84 *linear code*

A *linear code* of length n over $\mathbf{F}_q$ is a subspace of $\mathbf{F}_q^n$. The *dual code* $C^\perp$ of C is the orthogonal complement of the subspace C of $\mathbf{F}_q^n$. The *dimension* of the linear code C is the dimentions of C as a vector space over $\mathbf{F}_q$, that is to say, $\dim C$. A linear code C of length n and dimension k over $\mathbf{F}_q^n$ is called a q-ary $[n,k]$-code, or an (n,q^k)-linear code. If the distance d of C is known, it is called an $[n,k,d]$-linear code. Furthermore, C is said to be *self-orthogonal* if $C \subseteq C^\perp$, and *self-dual* if $C = C^\perp$.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 85 *Hamming weight*

Let $\mathbf{x}$ be a word in $\mathbf{F}_q^n$. Then, the *Hamming weight* $w(\mathbf{x})$ of $\mathbf{x}$ is defined as the number of non-zero letters in $\mathbf{x}$. In other words, $w(\mathbf{x}) = d(\mathbf{x},\mathbf{0})$, where $\mathbf{0}$ is the zero word and $d(\mathbf{x},\mathbf{y})$ is the Hamming distance between two words $\mathbf{x}$ and $\mathbf{y}$ in $\mathbf{F}_q^n$. For each element x of $\mathbf{F}_q$, the Hamming weight may be defined as

$$w(x) = d(x,0) = \begin{cases} 1, & \text{if } x \neq 0 \\ 0, & \text{if } x = 0 \end{cases}$$

Then for $\mathbf{x} = (x_1, \ldots, x_n)$ in $\mathbf{F}_q^n$,

$$w(\mathbf{x}) = w(x_1) + \cdots + w(x_n)$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 57 *Hamming weight and distance*

Let $\mathbf{x}$ and $\mathbf{y}$ be two words in $\mathbf{F}_q^n$. Then $d(\mathbf{x}, \mathbf{y}) = w(\mathbf{x} - \mathbf{y})$.

Proof. For each pair of letters x and y in $\mathbf{F}_q$, we know that $d(x, y) = 0$ if and only if $x = y$, that is if and only if $x - y = 0$, or equivalently $w(x - y) = 0$. The proof follows since $w(\mathbf{x}) = \sum_{i=1}^{n} w(x_i)$ and $d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} d(x_i, y_i)$. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Corollary 57[1] *q even*

Let q be an even positive integer. Then, for any two words $\mathbf{x}$ and $\mathbf{y}$ in $\mathbf{F}_q^n$ we have $d(\mathbf{x}, \mathbf{y}) = w(\mathbf{x} + \mathbf{y})$.

Proof. The proof follows from the fact that $a = -a$ for all a in $\mathbf{F}_q$ when q is even. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Theorem 58 *inequality*

Let $\mathbf{x}$ and $\mathbf{y}$ be two words in $\mathbf{F}_2^n$. Then, $w(\mathbf{x}) + w(\mathbf{y}) \geq w(\mathbf{x} + \mathbf{y})$.

Proof. For $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ in $\mathbf{F}_q^n$, let

$$\mathbf{x} * \mathbf{y} = (x_1 y_1, \ldots, x_n y_n)$$

Then, for $q = 2$ and $n = 1$,

x	y	$x * y$	$w(x) + w(y) - 2w(x * y)$	w(x+y)
0	0	0	0	0
0	1	0	1	1
1	0	0	1	1
1	1	1	0	0

From this together with Definition 85 we know that

$$w(\mathbf{x} + \mathbf{y}) = w(\mathbf{x}) + w(\mathbf{y}) - 2w(\mathbf{x} * \mathbf{y})$$

for $\mathbf{x}$ and $\mathbf{y}$ in $\mathbf{F}_2$, and thus the proof is implied. ¶

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Problem 10

Prove for any prime power q and $\mathbf{x}, \mathbf{y}$ in $\mathbf{F}_q^n$, that

$$w(\mathbf{x}) + w(\mathbf{y}) \geq w(\mathbf{x} + \mathbf{y}) \geq w(\mathbf{x}) - w(\mathbf{y})$$

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 86 *elementary row operation*

Let A be a matrix over $\mathbf{F}_q$. An *elementary row operation* performed on A is any one among the following.

a. interchange of two rows

b. multiplication of a row by a non-zero scalar

c. replacement of a row by its summation with a scalar multiple of another row

Two matrices are said to be *row equivalent* to each other if one is obtainable from another by a sequence of elementary row operations.

Kit Tyabandha, PhD Department of Mathematics, Mahidol University

Definition 87 *equivalent matrices*

Any matrix is row equivalent to a matrix in *row echelon* (RE) form or *reduced row echelon* (RRE)† form formed by a sequence of elementary row operations done upon itself. The RRE form of any given matrix is unique, but its RE's may not be so.

Appendix

Course Outline

Week	*Date*	*Topic of lecture*	*Hours*
1	28 Oct 2005	Error and distance	3
2	4 Nov 2005	Entropy and mutual information	3
3	11 Nov 2005	Group, field and finite field	3
4	18 Nov 2005	Bounds in coding	3
5	25 Nov 2005	Group, polynomial & Hamming codes	3
6	2 Dec 2005	Finite field- and BCH codes	3
7	9 Dec 2005	Linear codes	3
8	6 Jan 2006	Cyclic codes	3
9	13 Jan 2006	Goppa codes	3
10	20 Jan 2006	MDS code	3
11	10 Feb 2006	Cryptography	3

Quiz 1
Coding Theory

20^{th} January 2006

Time: 1 hours (12:30–1:30pm)

1. Write the addition [3]† and multiplication [4] tables for $\mathbf{Z}_6$.

Solution. The addition table,

+	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

#

The multiplication table,

·	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

#

2. Given ISBN 0 19 8538□3 0. Find the missing digit □.[3]

Solution. For ISBN $x_1 \ldots x_1 0$,

$$\sum_{i=1}^{10} i x_i \equiv (\mathrm{mod}\, 11)$$

Writing y for □,

$$0+1(2)+9(3)+8(4)+5(5)+3(6)+8(7)+y(8)+3(9) = 187+8y \equiv 0(\mathrm{mod}\, 11)$$

Hence $y = 0$, and the ISBN is therefore 0 19 853803 0.

#

† Numbers between square brackets are marks.

3. Let $f(x) = 1 + x^2 + x^3$. Show whether $f(x)$ is irreducible over $\mathbf{Z}_2$.[4] Then find $\mathbf{Z}_2[x]/(f(x))$.[4] And then draw the addition [5] and multiplication [7] tables of $\mathbf{Z}_2[x]/(f(x))$.

Solution. We note that $f(x)$ is of degree 3. Suppose $f(x)$ be reducible. Then it would have a linear factor x or $1+x$, which would make 0 and 1 roots of $f(x)$. But $g(0) = g(1) = 1$, which is in $\mathbf{Z}_2$. Therefore $f(x)$ is irreducible.

#

$$\mathbf{Z}_2[x]/\left(1 + x^2 + x^3\right) = \left\{0, 1, x, 1+x, x^2, x+x^2, 1+x^2, 1+x+x^2\right\}$$

#

The addition table,

$+$	0	1	x	$1+x$	x^2	$x+x^2$	$1+x^2$	$1+x+x^2$
0	0	1	x	$1+x$	x^2	$x+x^2$	$1+x^2$	$1+x+x^2$
1	1	0	$1+x$	x	$1+x^2$	$1+x+x^2$	x^2	$x+x^2$
x	x	$1+x$	0	1	$x+x^2$	x^2	$1+x+x^2$	$1+x^2$
$1+x$	$1+x$	x	1	0	$1+x+x^2$	$1+x^2$	$x+x^2$	x^2
x^2	x^2	$1+x^2$	$x+x^2$	$1+x+x^2$	0	x	1	$1+x$
$x+x^2$	$x+x^2$	$1+x+x^2$	x^2	$1+x^2$	x	0	$1+x$	1
$1+x^2$	$1+x^2$	x^2	$1+x+x^2$	$x+x^2$	1	$1+x$	0	x
$1+x+x^2$	$1+x+x^2$	$x+x^2$	$1+x^2$	x^2	$1+x$	1	x	0

#

The multiplication table,

$\cdot$	0	1	x	$1+x$	x^2	$x+x^2$	$1+x^2$	$1+x+x^2$
0	0	0	0	0	0	0	0	0
1	0	1	x	$1+x$	x^2	$x+x^2$	$1+x^2$	$1+x+x^2$
x	0	x	x^2	$x+x^2$	$1+x^2$	1	$1+x+x^2$	$1+x$
$1+x$	0	$1+x$	$x+x^2$	$1+x^2$	1	$1+x+x^2$	x	x^2
x^2	0	x^2	$1+x^2$	1	$1+x+x^2$	x	$1+x$	$x+x^2$
$x+x^2$	0	$x+x^2$	1	$1+x+x^2$	x	$1+x$	x^2	$1+x^2$
$1+x^2$	0	$1+x^2$	$1+x+x^2$	x	$1+x$	x^2	$x+x^2$	1
$1+x+x^2$	0	$1+x+x^2$	$1+x$	x^2	$x+x^2$	$1+x^2$	1	$1+x$

#

Midterm Examination

Coding Theory

27^{th} January 2006

Time: 2 hours (12:30–2:30pm)

1. Our task is to design a Hamming code in the case where there are 4 check digits. First find the binary-representative matrix M to be used for the purpose.[1]† Find the length of the code word [1] and that of a message word [1]. What is the redundancy of this code?[1] Then construct the code.[5] And then code the message words 10101010101 [3] and 00111100101 [3].

Solution. We have $r = 4$, $n = 2^4 - 1 = 15$ and $m = 2^4 - 4 - 1 = 11$. Each code word has 15 digits and 4 check digits.

#

Each message The redundancy of the code is 4.

#

† Numbers placed between square brackets are marks.

check digits	b_1	b_2		b_4				b_8						
	$\uparrow$	$\uparrow$		$\uparrow$				$\uparrow$						
code word	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}	b_{14}	b_{15}
			$\downarrow$		$\downarrow$	$\downarrow$	$\downarrow$		$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
message word			a_1		a_2	a_3	a_4		a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}

The $(2^r - 1) \times r$ matrix M is then

$$M = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

Then form the matrix equation

$$(\, b_1 \quad b_2 \quad \cdots \quad b_{15} \,)\, M = \mathbf{0}$$

which gives four equations in four unknown,

$$\begin{aligned} b_8 + b_9 + b_{10} + b_{11} + b_{12} + b_{13} + b_{14} + b_{15} &= 0 \\ b_4 + b_5 + b_6 + b_7 + b_{12} + b_{13} + b_{14} + b_{15} &= 0 \\ b_2 + b_3 + b_6 + b_7 + b_{10} + b_{11} + b_{14} + b_{15} &= 0 \\ b_1 + b_3 + b_5 + b_7 + b_9 + b_{11} + b_{13} + b_{15} &= 0 \end{aligned}$$

The message word 10101010101 becomes the code word

$$b_1 \quad b_2 \quad 1 \quad b_4 \quad 0 \quad 1 \quad 0 \quad b_8 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1$$

Hence, $b_8 = 0$, $b_4 = 1$, $b_2 = 0$ and $b_1 = 1$, and the code word is

101101001010101

#

The message 00111100101 becomes the code word

$$b_1 \quad b_2 \quad 0 \quad b_4 \quad 0 \quad 1 \quad 1 \quad b_8 \quad 1 \quad 1 \quad 0 \quad 0 \quad 1 \quad 0 \quad 1$$

which gives us $b_8 = 0$, $b_4 = 0$, $b_2 = 0$ and $b_1 = 1$. Therefore the code word is 100001101100101.

#

2. Show that the polynomial x^3+x^2+1 is irreducible [4] and the element $\alpha = x + \langle x^3+x^2+1\rangle$ of $F[x]/\langle x^3+x^2+1\rangle$ primitive [4]. Then use this to construct a binary BCH code of length 7 and minimum distance 3.[5] And then find the code word for the message 1101.[2]

Solution. Suppose x^3+x^2+1 is reducible, then it must have either x or $x+1$ as a factor, then $x=0$ or 1 is a root of x^3+x^2+1. But

$$\begin{array}{r|l} x & x^3+x^2+1 \\ \hline & x^3 \\ \hline & x^2+1 \\ & x^2 \\ \hline & 1 \end{array} \qquad \begin{array}{r|l} x+1 & x^3+x^2+1 \\ \hline & x^3+x^2 \\ \hline & 1 \end{array}$$

from which we can see that both $x|x^3+x^2+1$ and $x+1|x^3+x^2+1$ give a remainder 1. Therefore neither x nor $x+1$ divides x^3+x^2+1, hence the latter is irreducible.

#

We know that $f(x)$ in $F_p[x]$ of degree n is primitive if $f(x)|x^{p^n-1}-1$ and $f(x) \not| x^k-1$ for any $k < p^n-1$. We have $x^{2^3-1}-1 = x^7-1$. Then

$$\begin{array}{r|l} x^3+x^2+1 & x^4+x^3+x^2+1 \\ \hline & x^7-1 \\ & x^7+x^6+x^4 \\ \hline & x^6+x^4+1 \\ & x^6+x^5+x^3 \\ \hline & x^5+x^4+x^3+1 \\ & x^5+x^4+x^2 \\ \hline & x^3+x^2+1 \\ & \to 0 \end{array} \quad \Rightarrow \quad x^3+x^2+1|x^7-1$$

For $k<7$; if $k=6$;

$$\begin{array}{r|l} x^3+x^2+1 & x^3+x^2+x \\ \hline & x^6-1 \\ & x^6+x^5+x^3 \\ \hline & x^5+x^3+1 \\ & x^5+x^4+x^2 \\ \hline & x^4+x^3+x^2+1 \\ & x^4+x^3+x \\ \hline & x^2+x+1 \\ & \neq 0 \end{array} \quad \Rightarrow \quad x^3+x^2+1 \not| x^6-1$$

If $k=5$;

$$
\begin{array}{r|l}
x^3+x^2+1 & x^2+x+1 \\
\hline
 & x^5-1 \\
 & x^5+x^4+x^2 \\
\hline
 & x^4+x^2+1 \\
 & x^4+x^3+x \\
\hline
 & x^3+x^2+x+1 \\
 & x^3+x^2+1 \\
\hline
 & x \qquad \neq 0 \quad \Rightarrow \quad x^3+x^2+1 \not| x^5-1
\end{array}
$$

If $k=4$;

$$
\begin{array}{r|l}
x^3+x^2+1 & x+1 \\
\hline
 & x^4-1 \\
 & x^4+x^3+x \\
\hline
 & x^3+x+1 \\
 & x^3+x^2+1 \\
\hline
 & x^2+4 \quad \neq 0 \quad \Rightarrow \quad x^3+x^2+1 \not| x^4-1
\end{array}
$$

When $k=3$, $x^3+x^2+1 \not| x^3-1$ is obvious. Therefore $\alpha = x + \langle x^3+x^2+1 \rangle$ is a primitive of $F[x]/\langle x^3+x^2+1 \rangle$.

Then, α satisfies $\alpha^3+\alpha^2+1=0$. For F a finite field of order p^n with k as its prime subfield, we know that α and α^p have the same minimum polynomial over k for every $\alpha \in F$. Here $p=1$; therefore α and α^2 have the same minimum polynomial, hence the generating polynomial is x^3+x^2+1. Let the message be $a_0a_1a_2a_3$. Then the message polynomial is $a(x)=a_0+a_1x+a_2x^2+a_3x^3$. The corresponding code polynomial is therefore $a(x)\left(x^3+x^2+1\right)$. In other words,

$$a_0+a_1x+(a_0+a_2)\,x^2+(a_0+a_1+a_3)\,x^3+(a_1+a_2)\,x^4+(a_2+a_3)\,x^5+a_3x^6$$

The code word is thus

$$(a_0, a_1, (a_0+a_2), (a_0+a_1+a_3), (a_1+a_2), (a_2+a_3), a_3)$$

#

For the message 1101, the code word is 1111111.

#

Quiz 2
Coding Theory

3^{rd} February 2006

Time: 1 hours (12:30–1:30pm)

1. Let $S = \{11010, 10111, 01010, 01101\}$. Find a basis for $C\langle S\rangle$. [5] What is the dimension k? [1] Find the binary code C. [4] Find also all cosets of C. [10]

Solution. Form and reduce our matrix A,

$$A = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Hence the basis is $\{11010, 01101, 00111\}$

#

The basis has three components, therefore dimension k is 3.

#

The binary code is

$$C = \{00000, 11010, 01101, 00111, 10111, 11101, 01010, 10000\}$$

#

	cofactor	$\to$	*words*
I	$00000 + C$	$\to$	00000, 11010, 01101, 00111, 10111, 11101, 01010, 10000
II	$00001 + C$	$\to$	00001, 11010, 01100, 00110, 10110, 11100, 01011, 10001
III	$00010 + C$	$\to$	00010, 11000, 01111, 00101, 10101, 11111, 01000, 10010
IV	$00100 + C$	$\to$	00100, 11110, 01001, 00011, 10011, 11001, 01110, 10100

Since there other coset leaders all give one of these four cosets, therefore the number of cosets is four and all of them are listed above.

#

2. Based on te factorisation $x^6 - 1 = (1 + x)^2 \left(1 + x + x^2\right)^2$, find a binary $[6, 3]$ cyclic code. [10]

Solution. Here we are given $k = 3$ and $n = 6$. List all nine monic divisors of $x^6 - 1$ and note the degree k, that is the number of different bases, of each. This is $x^i(\cdot)$, $i < k$.

$$\begin{array}{lcl} 1 & \rightarrow & k=6 \\ 1+x & \rightarrow & k=6 \\ 1+x+x^2 & \rightarrow & k=5 \\ (1+x)^2 & \rightarrow & k=4 \\ (1+x)\left(1+x+x^2\right) & \rightarrow & k=3 \\ (1+x)^2\left(1+x+x^2\right) & \rightarrow & k=2 \\ \left(1+x+x^2\right)^2 & \rightarrow & k=2 \\ (1+x)\left(1+x+x^2\right) & \rightarrow & k=1 \\ \left(1+x^6\right) & \rightarrow & k=0 \end{array}$$

For $k=3$;

$$(1+x)\left(1+x+x^2\right) = 1+x^3$$

$$\begin{array}{lcl} 0\cdot\left(1+x^3\right) & \rightarrow & 000000 \\ 1\cdot\left(1+x^3\right) & \rightarrow & 100100 \\ x\cdot\left(1+x^3\right) & \rightarrow & 010010 \\ x^2\cdot\left(1+x^3\right) & \rightarrow & 001001 \end{array}$$

And the pairwise additions among these give us the remaining code words. Then,

$$C = \{000000, 100100, 010010, 001001, 110110, 101101, 011011, 111111\}$$

#

Quiz 3
Coding Theory

10^{th} February 2006

Time: 1 hour (12:30–1:30pm)

1. Consider the matrix

$$A = \begin{bmatrix} 1 & 3 & 4 \\ 2 & 3 & 1 \\ 2 & 4 & 4 \end{bmatrix}$$

over $GF(5)$. Show that A can give minimum distance separable (MDS) codes. [6] Find two such codes if they exist, and give either a generator matrix or a parity check matrix for each of them. [6] Then give the code words and encoding functions for each. [8]

Solution. Examine the values of determinant for all submatrices of A, $\begin{vmatrix} 1 & 3 \\ 2 & 3 \end{vmatrix} = 2$, $\begin{vmatrix} 1 & 4 \\ 2 & 1 \end{vmatrix} = 3$, $\begin{vmatrix} 3 & 4 \\ 3 & 1 \end{vmatrix} = 1$, $\begin{vmatrix} 1 & 3 \\ 2 & 4 \end{vmatrix} = 3$, $\begin{vmatrix} 1 & 4 \\ 2 & 4 \end{vmatrix} = 1$, $\begin{vmatrix} 3 & 4 \\ 4 & 4 \end{vmatrix} = 1$, $\begin{vmatrix} 2 & 3 \\ 2 & 4 \end{vmatrix} = 2$, $\begin{vmatrix} 2 & 1 \\ 2 & 4 \end{vmatrix} = 1$, $\begin{vmatrix} 3 & 1 \\ 4 & 4 \end{vmatrix} = 3$, $\begin{vmatrix} 1 & 3 & 4 \\ 2 & 3 & 1 \\ 2 & 4 & 4 \end{vmatrix} = 3$. We can see that no submatrices of A are singular, therefore we may obtain from A two MDS codes, namely the $[6, 3, -]$ code over $GF(5)$ with the generator matrix $G = (I_3 \quad A)$ and the $[6, 3, -]$ code over $GF(5)$ with the parity check matrix $H - (A \quad I_3)$. For the first one, the generating function is

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 & 3 & 4 \\ 0 & 1 & 0 & 2 & 3 & 1 \\ 0 & 0 & 1 & 2 & 4 & 4 \end{pmatrix}$$

Then,

$$(\, a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5 \quad a_6 \,) = (\, a_1 \quad a_2 \quad a_3 \,) \begin{pmatrix} 1 & 0 & 0 & 1 & 3 & 4 \\ 0 & 1 & 0 & 2 & 3 & 1 \\ 0 & 0 & 1 & 2 & 4 & 4 \end{pmatrix}$$

and the encoding functions becomes

$$\begin{aligned} a_4 &= a_1 + 2a_2 + 2a_3 \\ a_5 &= 3a_1 + 3a_2 + 4a_3 \\ a_6 &= 4a_1 + a_2 + 4a_3 \end{aligned}$$

#

The code words are $C = \{100134; 010231; 001244; 110310; 101323; 11420\}$.

#

For the $[6,3,-]$ code, from the parity check matrix we have the generating function

$$G = (\, I_3 \quad -A^T \,) = \begin{pmatrix} 1 & 0 & 0 & 4 & 3 & 3 \\ 0 & 1 & 0 & 2 & 2 & 1 \\ 0 & 0 & 1 & 1 & 4 & 1 \end{pmatrix}$$

Then,

$$(\, a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5 \quad a_6 \,) = (\, a_1 \quad a_2 \quad a_3 \,) \begin{pmatrix} 1 & 0 & 0 & 4 & 3 & 3 \\ 0 & 1 & 0 & 2 & 2 & 1 \\ 0 & 0 & 1 & 1 & 4 & 1 \end{pmatrix}$$

and the encoding functions become

$$\begin{aligned} a_4 &= 4a_1 + 2a_2 + a_3 \\ a_5 &= 3a_1 + 2a_2 + 4a_3 \\ a_6 &= 3a_1 + a_2 + a_3 \end{aligned}$$

#

And then the code is $C = \{100433, 010221, 001141, 110104, 101024, 011312\}$.

#

Final Examination
Coding Theory

23^{rd} February 2006

Time: 3 hours (12:30–3:30pm)

1. A channel matrix P is a matrix whose elements p_{ij} has the value of the probability that an input a_i in an input alphabet Σ_1 will produce an output b_j in an output alphabet Σ_2. The r^{th} *extension* of this channel is the discrete memoryless channel with input alphabet $\Sigma_1^{(r)} = a_1 a_2 \dots a_r$, output alphabet $\Sigma_2^{(r)} = b_1 b_2 \dots b_r$, and channel matrix $P^{(r)}$ whose components are $p_{ij}^{(r)} = p(b_1|a_1)\, p(b_2|a_2) \cdots p(b_r|a_r)$. Furthermore, P is said to be a *stochastic matrix* if $p_{ij} \geq 0$ for all i, j and all row sums are equal to 1, that is to say,

$$\sum_j p_{ij} = 1$$

a. Consider a binary symmetric channel, that is to say, a mapping from input to output such that $\Sigma_1 = \Sigma_2 = \{0, 1\}$. The second extension of this channel has input and output alphabet $\{00, 01, 10, 11\}$. If the channel matrix for this binary channel is

$$p = \begin{bmatrix} q & p \\ p & q \end{bmatrix}$$

what would be the channel matrix for the second extension of this channel? [4] Find $P^{(2)}$, if it is stochastic. [2]

b. Next, consider the binary erasure channel whose input alphabet is $\Sigma_1 = \{0, 1\}$ and output alphabet $\Sigma_2 = \{0, 1, *\}$. The channel matrix in this case may be described as

$$P = \begin{bmatrix} 1-\varepsilon & 0 & \varepsilon \\ 0 & 1-\varepsilon & \varepsilon \end{bmatrix}$$

Find the channel matrix for the second extension of this channel. [9]

2.

a. Explain *uncertainty* and *entropy*, *conditional entropy*, and *mutual information*, giving examples where appropriate. [2]

b. Let $H(X)$ be the entropy of a random variable X that takes on a finite set of values with probabilities $p_1, p_2, \dots, p_n$, that is $H(X) = -\sum_k p_k \ln p_k$. Show how the following statements are true. [8]

   i. $H(p_1, \dots, p_n)$ is at maximum when $p_1 = p_2 = \dots = p_n = \frac{1}{n}$

   ii. $H(p_1, \dots, p_n) = H\left(p_{\pi(1)}, \dots, p_{\pi(n)}\right)$

   iii. $H(p_1, \dots, p_n) \geq 0$, where the equality holds only when one of the p_i's is 1.

iv. $H(p_1, \ldots, p_n, 0) = H(p_1, \ldots, p_n)$

v.
$$H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) \leq H\left(\frac{1}{n+1}, \frac{1}{n+1}, \ldots, \frac{1}{n+1}\right)$$

vi. $H(p_1, \ldots, p_n)$ is a continuous function of its arguments.

vii.
$$H\left(\frac{1}{mn}, \frac{1}{mn}, \ldots, \frac{1}{mn}\right) = H\left(\frac{1}{m}, \ldots, \frac{1}{m}\right) + H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$$

viii. Let $p = p_1 + \cdots + p_m$ and $q = q_1 + \cdots + q_n$, where $p + q = 1$, p and q are positive and p_i and p_j non-negative. Then,

$$H(p_1, \ldots, p_m, q_1, \ldots, q_n) =$$

$$H(p, q) + pH(p, q) + pH\left(\frac{p_1}{p}, \ldots, \frac{p_m}{p}\right) + qH\left(\frac{q_1}{q}, \ldots, \frac{q_n}{q}\right)$$

3.

a. Give the meaning and explanation of the terms *cryptography*, *cryptanalysis*, *cryptology*, *key*, *plain text*, *code*, *cipher*, *encode*, *decode*, *encrypt*, *decrypt* and *nondeterministic polynomial–complete*. [6]

b. Caesar cipher shifts all message letters three positions to the right on the ordered list of characters of the roman alphabet. Encode the previous sentence using Caesar cipher. [3]

c. One of the most basic encryption algorithms is the transposition of order d, in which the message is divided into blocks of length d and then a permutation π of $1, 2, \ldots, d$ is applied to each block. As an example, when $d = 5$ and $\pi = \{3\,4\,5\,2\,1\}$, Shakespeare's 'Fair is foul, and foul is fair' is encrypted into 'rfai fis a ul,siul ri fa'. Again, let $d = 5$ and $\pi = \{5\,3\,1\,2\,4\}$, then try to decrypt, 'thi w cyrms-sbo iw ooth se hatatbrls s o'. Find what the original message is. [6]

4. Write an essay of approximately 25 lines on either one of the following topics: group, polynomial, Hamming, linear, or cyclic codes, MDS, Goppa, Hadamard, or quadratic residue code, automorphism group of a code, entropy and mutual information, or cryptography. [10]

Students' scores
Coding Theory
2005–6
14^{th} January, 2007

Quiz 1

Quiz 1 was done on 20 January 2006. There were three questions

Name	*ID*	*Question*			*Total*
		1	2	3	
Kạṇṭaditạ	4505016	7	3	5	15
Jȧrañyạ	4505039	7	0.2	12.1	19.3
Ṇaṭhȧbol	4505054	7	3	4.7	14.7
Dhȧṇeṣ́	4505071	7	1	3.4	11.4
Rattiya	4505181	7	3	4.7	14.7
Rattayaporn	4505183	7	2	4.9	13.9
Vaurȧṇan	4505188	7	2	4.8	13.8
Saniwan	4505216	7	1	4.2	12.2
Vạsȧṇạ	4505204	7	2	4.9	13.9
Śiridibẏa	4505220	7	0	4.9	11.9

Table 6 *Students' midterm scores.*

The scores and ranks for Quiz 1 are shown in Table 7.

ID	*score*	*rank*	*ID*	*score*	*rank*	*ID*	*score*	*rank*
45016	15	2	45071	11.4	8	45188	13.8	5
45039	19.3	1	45181	14.7	3	45204	13.9	4
45054	14.7	3	45183	13.9	4	45216	12.2	6
						45220	11.9	7

Table 7 *Mark and rank of students' scores from Quiz 1.*

The total final score for Quiz 1 is 10. The mean is 7.04, median 6.95, minimum 5.70 and Maximum 9.65. The standard deviation is 1.11. Figure 8 shows the distribution.

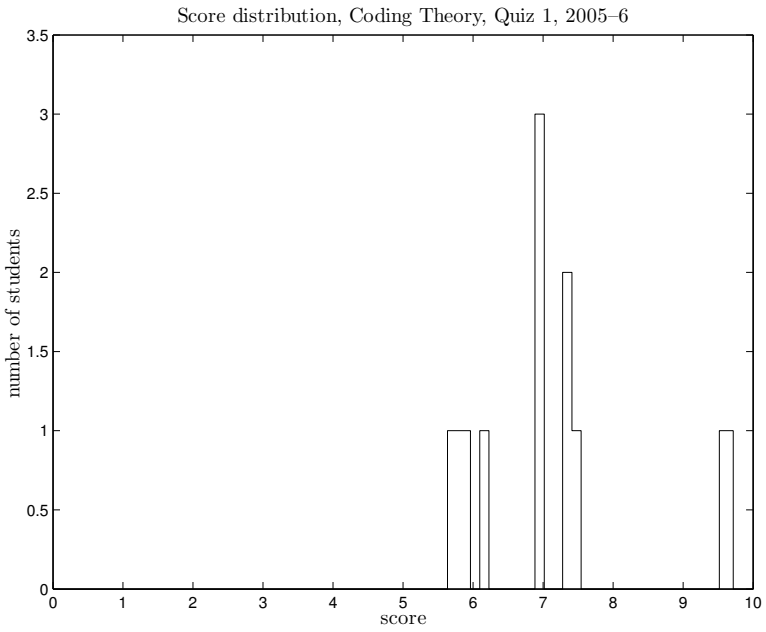


Figure 8 *Distribution of students' Quiz 1 scores.*

Midterm Exam

Midterm Exam was done on 27 January 2006. There were two questions of 15 points each. The total collected mark was 20. Table 8 gives scores for each question.

ID	*M1(15)*	*M2(15)*
45016	14.5	15
45039	15	3.8
45054	15	11.2
45071	14.5	4
45092	15	10
45181	15	14.7
45183	15	15
45188	14.1	15
45204	15	15
45216	15	3.8
45220	14.3	11.4

Table 8 *Mark and rank of students' scores from Midterm Exam.*

The total final score for the midterm exam is 20. Of this, the mean is 17.05, median 17.47, minimum 12.33 and maximum 20. The standard deviation is 3.17. The scaled scores are given in Table 9.

ID	*score*	*rank*
45016	19.67	3
45039	12.53	8
45054	17.47	5
45071	12.33	9
45092	16.67	7
45181	19.80	2
45183	20	1
45188	19.40	4
45204	20	1
45216	12.53	8
45220	17.13	6

Table 9 *Mark and rank of students' scores from Midterm Exam.*

Figure 9 shows the distribution of midterm scores.

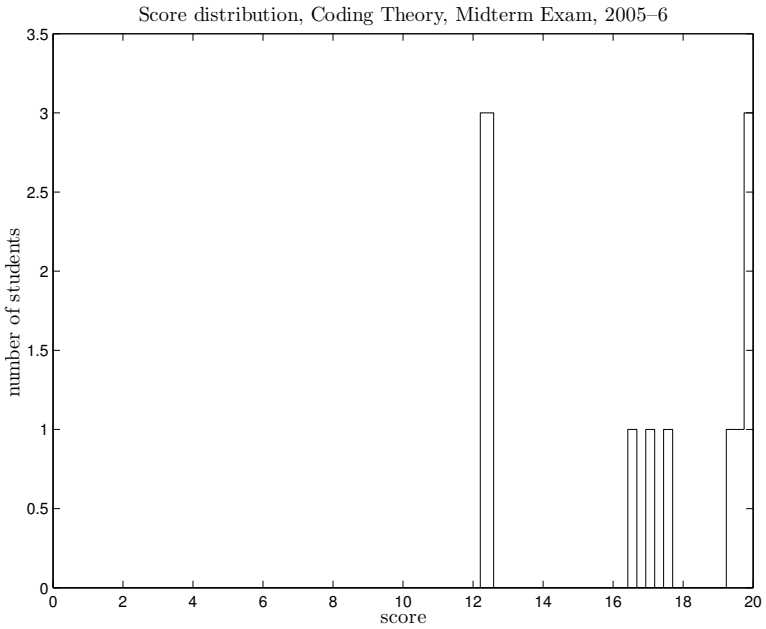


Figure 9 *Distribution of students' Midterm scores.*

Quiz 2

Quiz 2 was done on 3 February 2006. There were two questions. The total mark is 30, which is later scaled down to 10. Table 10 gives the scores for each question.

ID	*score*	*rank*
45016	11.5	10
45039	0.5	2.5
45054	19	9.8
45071	6	1.3
45092	8	6.3
45181	8	5.7
45183	8	1
45188	4.8	7
45204	1	7.1
45216	8	5
45220	7.3	1.51

Table 10 *Mark and rank of students' scores from Quiz 2.*

Figure 10 shows the distribution of Quiz 2 scores.

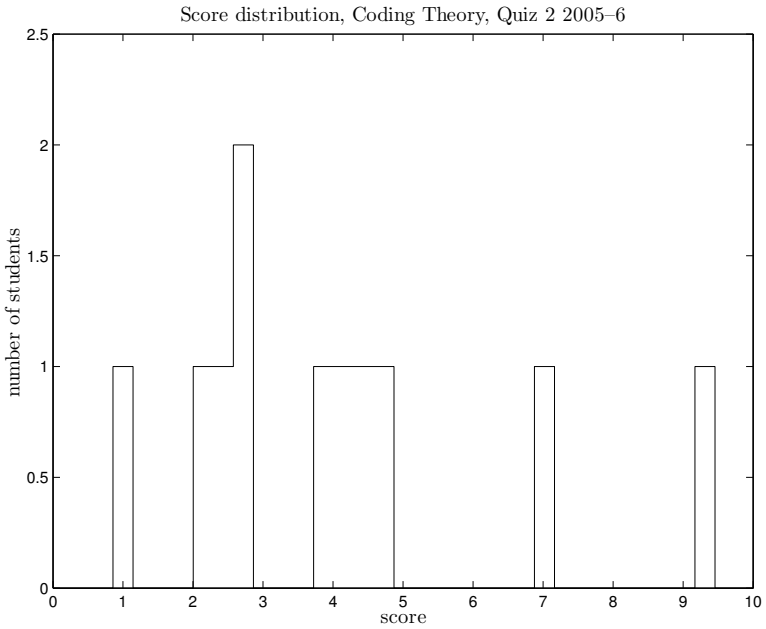


Figure 10 *Distribution of students' Quiz 2 scores.*

For Quiz 2 the scaled total score is 10. Then the mean after scaling is 4.22, median 3.93, minimum 1, maximum 9.60, and the standard deviation 2.39. Table 11 gives the scaled score together with ranking.

ID	*score*	*rank*
45016	7.17	2
45039	1	11
45054	9.60	1
45071	2.43	10
45092	4.77	3
45181	4.57	4
45183	3	7
45188	3.93	6
45204	2.70	9
45216	4.33	5
45220	2.93	8

Table 11 *Mark and rank of students' scores from .*

Quiz 3

Quiz 3 took place on 10 February 2006. There was only question the marks of which is 20. This is later scaled down to 10.

ID	*score*	*rank*	*ID*	*score*	*rank*	*ID*	*score*	*rank*			
45016	16.7	1	45071	16.7	1	45183	16.7	1			
45039	15.2	3	45092	16.7	1	45188	16.7	1			
45054	16.7	1	45181	16.7	1	45204	16.6	2			
									45216	16.7	1
									45220	11.8	4

Table 12 *Scores and ranks of Quiz 3.*

From Quiz 3 the scaled mean is 8.05, and the median 8.35, minimum 5.90, maximum 8.35 and standard deviation 0.75.

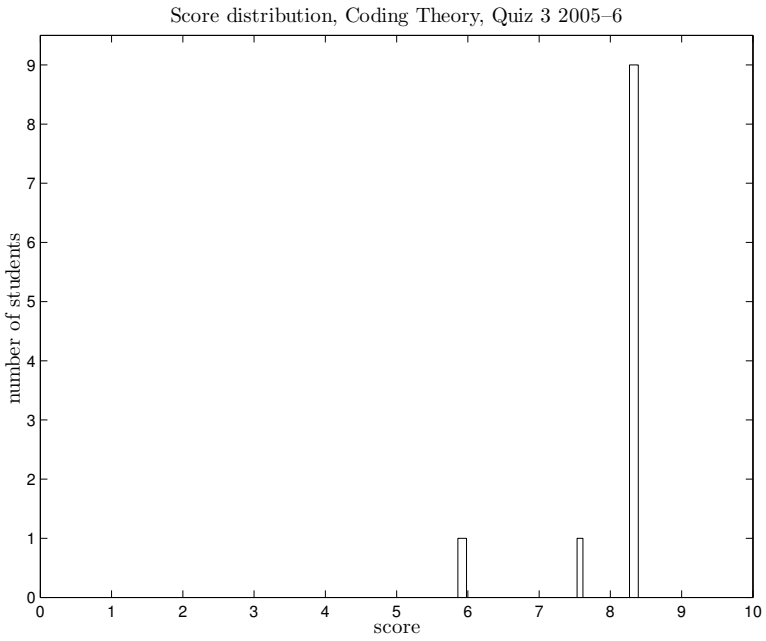


Figure 11 *Distribution of students' Quiz 3 scores.*

Final Exam

Final Exam was held on 24 February 2006. There were four questions, which add up to 50 marks in total. This is later scaled down to 30.

ID	*score*	*rank*	*ID*	*score*	*rank*	*ID*	*score*	*rank*			
45016	9.7	8	45071	10.9	6	45183	17.3	4			
45039	8.6	9	45092	19.4	2	45188	18.6	3			
45054	31	1	45181	9.9	7	45204	14.6	5			
									45216	7.9	10
									45220	7.3	11

Table 13 *Scores and ranks of Final Exam.*

The Final Exam scores are scaled from the original total of 50 into 30. Consequently they have as the mean 8.47, median 6.54, minimum 4.38, maximum 18.60 and standard deviation 4.27. Figure 12 gives the plot of the distribution of the scores.

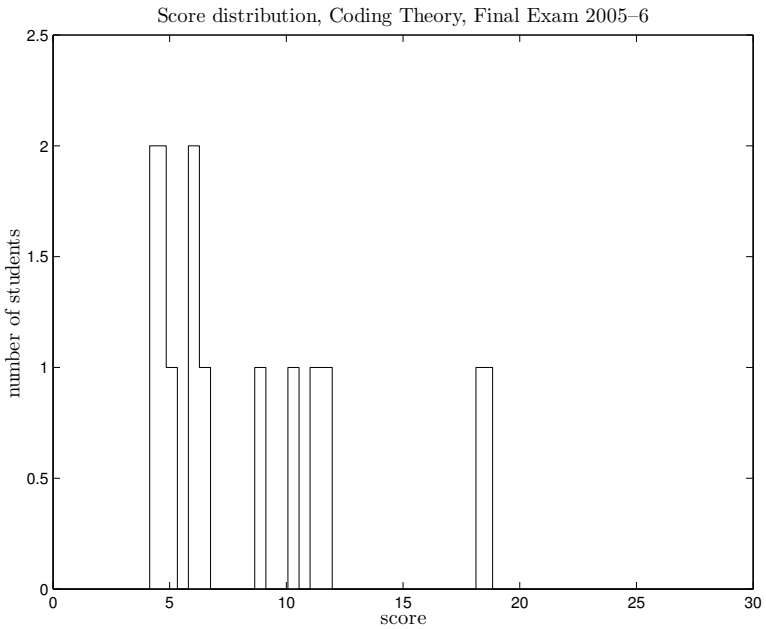


Figure 12 *Distribution of students' final exam scores.*

Practice and attendance

Then there are marks from practice and attendance. These are listed in Table 14.

ID	*Practice*	*Attendance*
45016	10	8.5
45039	9	8
45054	10	10
45071	10	10
45092	10	10
45181	10	9
45183	10	9.5
45188	10	10
45204	10	9.5
45216	10	10
45220	10	8.5

Table 14 *Practice and attendance.*

Total Scores

The total scores are given in Table 15.

ID	*Quiz 1* (10)	*Quiz 2* (10)	*Quiz 3* (10)	*Midterm* (20)	*Final* (30)	*Practice* (10)	*Attendance*
45016	7.5	7.17	8.35	19.67	5.82	8.5	10
45039	9.65	1	7.6	12.53	5.16	8	9
45054	7.35	9.6	8.35	17.47	18.6	10	10
45071	5.7	2.43	8.35	12.33	6.54	10	10
45092	–	4.77	8.35	16.67	11.64	10	10
45181	7.35	4.57	8.35	19.8	5.94	9	10
45183	6.95	3	8.35	20	10.38	9.5	10
45188	6.9	3.93	8.35	19.40	11.16	10	10
45204	6.95	2.7	8.3	20	8.76	9.5	10
45216	6.1	4.33	8.35	12.53	4.74	10	10
45220	5.95	2.93	5.9	17.13	4.38	8.5	10

Table 15 *Total score, Coding Theory, second term, 2005–6.*

The total score has as the mean 63.46, median 65.01, minimum 52.94, maximum 81.37 and standard deviation 8.46. The rank is shown together with the score for each student in Table 16.

ID	*score*	*rank*	*ID*	*score*	*rank*	*ID*	*score*	*rank*			
45016	67	4	45071	55.36	9	45183	68.18	3			
45039	52.94	11	45092	61.42	7	45188	69.74	2			
45054	81.37	1	45181	65.01	6	45204	66.21	5			
									45216	56.06	8
									45220	54.8	10

Table 16 *Total score and rank, Coding Theory, 2005–6.*

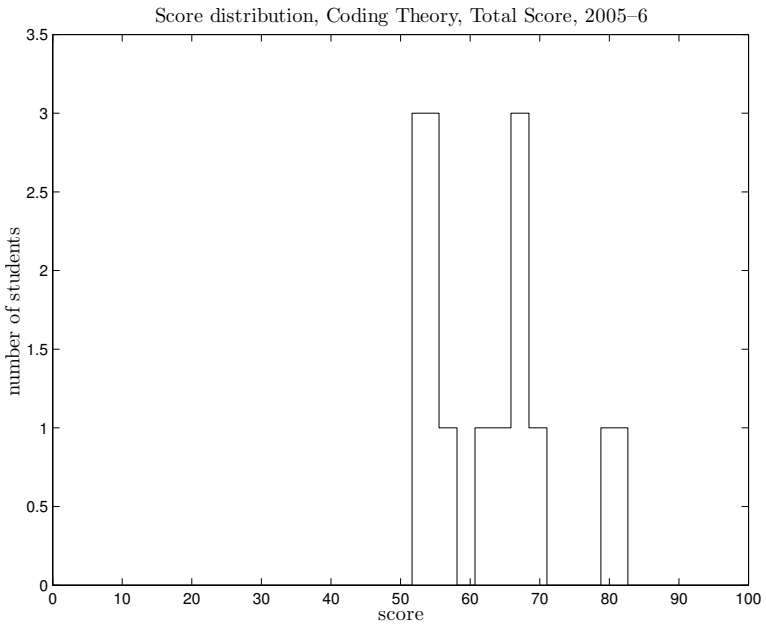


Figure 13 *Distribution of students' total score, Coding Theory.*

For grades, we look at two candidate grading scheme as shown in Table 17. Scheme A more closely resembles a hard grading scheme, that is to say, one which is independent of performance of student. Scheme B is attached to the range used in drawing the histogram of Figure 13. In that figure there appear three clusters of grades. These we make correspond to the three grades given, that is A, B^+ and B.

Grade	*Mark range*
A	$(75, 100]$
B^+	$(60, 75]$
B	$(50, 60]$

Table 17 *Grading schemes.*

According to these scheme, the grades are thus as shown in Table 18.

ID	*Score*	*Grade*
45016	67	B^+
45039	52.94	B
45054	81.37	A
45071	55.36	B
45092	61.42	B^+
45181	65.01	B^+
45183	68.18	B^+
45188	69.74	B^+
45204	66.21	B^+
45216	56.06	B
45220	54.8	B

Table 18 *Grades according to our grading scheme of Table 17.*

Kit Tyabandha
Mahidol, Bangkok
14^{th} January, 2007

Authors' profile

Kit Tyabandha
arthur.tyabandha@web.de

Education

2004	PhD	University of Manchester (UK)
1995	MSc	University of Manchester Institute of Science and Technology (UK)
1993	BEng	Electrical Engineering Chulalongkorn University (Thailand)
1992	BSc	Computer Science Ramkhamhaeng University (Thailand)
1991	BEng	Mineral Engineering Chulalongkorn University (Thailand)
1983	6^{th} Form	Ashburton College (New Zealand) Certificate

Other books published by Kittix Books

- Kittiśakd́xi Tiẏàbandẏha. Bhaṣạ Angkṛiṣ an nàsoncaị. 2000. ISBN 974-346-182-5
- Kittiśakd́xi Tiẏàbandẏha. Plạe kled́ Angkṛiṣ. 2000. ISBN 974-346-765-3
- Kittisak N Tiyapan. Voronoi Translated. Introduction to Voronoi tessellation and essays by G L Dirichlet and G F Voronoi. 2001. ISBN 974-13-1503-1
- Kit Nui Tiyapan. Percolation within percolation and Voronoi Tessellation. 2003. ISBN 974-91036-1-0
- Kit Tiyapan. Thai grammar, poetry and dictionary. 2003. ISBN 974-17-1861-6
- Kit Tiyapan. A Lanna in town. 2003. ISBN 974-17-1860-8
- Kit Tiyapan. A Kiwi Lanna. 2003. ISBN 974-91237-3-5
- Kit Tiyapan. A British Lanna. 2003. ISBN 974-91237-4-3
- Kit Tiyapan. (Kippu Chaban) Edokko no Lanna. 2003. ISBN 974-91341-9-2
- Kit Tiyapan. The Siamese Lanna. 2003. ISBN 974-91341-8-4
- Kit Tyabandha and K N Tiyapan. Percolation within percolation and Voronoi Tessellation, revised edition. 2005. ISBN 974-93037-5-X

Kinder is a german word that means 'children'. So the name of the place should more correctly be written 'Kinderscout'. But then again there are Kinder Downfall and Kinder Low. To do the same thing everywhere would probably result in something seemingly out of place sitting in an English context. Therefore the name is normally written 'Kinder Scout'.

Kit Tyabandha
Bangkok, April 2006